

Nordiske Studier i Leksikografi

13

Nordiske Studier i Leksikografi

13

Rapport fra 13. Konference om Leksikografi i Norden
København 19.-22. maj 2015

Redigeret af

Asgerd Gudiksen

Henrik Hovmark

Skrifter udgivet af Nordisk Forening for Leksikografi
Skrift nr. 14

I samarbejde med
Nordisk Forskningsinstitut, Københavns Universitet

København 2016

Nordiske Studier i Leksikografi 13

© Nordisk Forening for Leksikografi og forfatterne, 2016

Redaktion: Asgerd Gudiksen og Henrik Hovmark

Alle bidrag er fagfællebedømte

Sats: Henrik Hovmark

Bidragydere til sats: Bo Nissen Knudsen og Liisa Deth Theilgaard

Omslag: Tarm Bogtryk A/S

Tryk: Tarm Bogtryk A/S

ISBN: 978-87-992447-6-8

ISSN: 1894-4663 (trykt)

ISSN: 2246-7823 (online)

ISSN: 0803-9313 (Skrifter udgivet af Nordisk Forening for Leksikografi)

Udgivet med støtte fra:

Nordplus Nordiske Sprog



Nordisk Forskningsinstitut, Københavns Universitet



Indholdsfortegnelse

FORORD..... s. 9

Plenarforedrag

ESBEN ALFORT:

Ord og IT s. 13

ODDRUN GRØNVIK:

Vitskaplegheit og samfunnsrelevans
for store ordboksverk s. 27

Tema: Den digitale ordbogsbruger

MARGRETHE HEIDEMANN ANDERSEN, JØRGEN NØRBY JENSEN

& ANITA ÅGERUP JERVELUND:

Fordele og ulemper ved løbende opdatering af den
digitale Retskrivningsordbog s. 65

KRISTÍN BJARNADÓTTIR:

The Case for Normalization: Linking Lexicographic
Resources for Icelandic s. 79

TOVE BJØRNESET:

LEXIN i Norge – Hva sier brukerne? s. 89

ERIK BÄCKERUD:

Indexeringen av SAOB s. 101

ØYSTEIN EEK:

Ordnnett, forlaget og brukerne s. 111

LOUISE HALLSTRØM ABILDGAARD, AMALIE GLARGAARD

& CARSTEN HANSEN:

Facebook – et nyt brugersegment for Ordbog over
Dansk Talesprog s. 119

POUL HANSEN:

Hvordan har netbrugernes præference for forskellige
emnespecifikke ordbøger ændret sig i perioden 2006-2014? s. 131

HELGA HILMISDÓTTIR & NINA MARTOLA:

Flerspråkig lexikografi och obesläktade språk
– erfarenheter från finska Islex s. 143

MARIT HOVDENAK & INGUNN INDREBØ IMS:	
Bruk av digitale ordbøker i norsk skule	s. 155
ELLERT ÞÓR JÓHANNSSON & SIMONETTA BATTISTA:	
Ordbog over det norrøne prosasprog Online:	
struktur og bruk	s. 165
PATRICK LEROYER:	
Bruger- og ekspertinddragelse ved udarbejdelse	
af online (fag)ordbøger: det kooperative princip	
i leksikografien	s. 177
HENRIK LORENTZEN, NICOLAI HARTVIG SØRENSEN	
& LARS TRAP-JENSEN:	
Den Danske Ordbog som app – hvorfor og hvordan?	s. 191
CHRISTIAN-EMIL SMITH ORE:	
Gamle ordbøker og digitale utgaver	s. 203
CAROLINE SANDSTRÖM & CAROLA ÅKERLUND:	
Personbetecknande verbavledningar i Ordbok över	
Finlands svenska folkmål. En utvärdering av	
ordbokens sökfunktioner	s. 217
HENRIK KØHLER SIMONSEN:	
Den mobile ordbogsbruger	s. 227
THOMAS TROELSGÅRD & NICOLAI HARTVIG SØRENSEN:	
Nyt liv til støvede ordbøger	s. 239
SEÁN D VRIELAND:	
Stamtræer og digitale etymologiske ordbøger	s. 251

Generelt program

TORBEN ARBOE:	
Konventionelle figurative enheder (idiomer) i	
Jysk Ordbog og andre større ordbøger	s. 263
STURLA BERG-OLSEN & DANIEL GUSFRE IMS:	
Dansk og norsk i norske ordbøker	s. 277
LARS BRINK:	
Synonymitis	s. 291
KEN FARØ:	
Inventariografi og andre nyere tilgange til leksikografi	s. 301
RUTH VATVEDT FJELD & ARASH SAIDI:	
Hvordan slippe inn i Platons hage? – Om kartlegging	
og dokumentasjon av norsk akademisk vokabular	s. 313

AURELIJA GRIŠKEVIČIENĖ & STURLA BERG-OLSEN: Å gå seg vill i en ordbok: <i>seg</i> og <i>V+seg</i> i tospråklige ordbøker	s. 327
JON HELGASON & EMMA SKÖLDBERG: Kodifisering av multietnisk ungdomsspråk	s. 337
LOUISE HOLMER, SVEN-GÖRAN MALMGREN & MONICA VON MARTENS: SAOLhist.se – för allmänt och vetenskapligt bruk	s. 349
HÅKAN JANSSON: Från ”tandborstord” till ”memilord” – om nyord och deras belägg	s. 359
HALLDÓRA JÓNSDÓTTIR & ÞÓRDÍS ÚLFARSDÓTTIR: Processering af en synonym-komponent i en flersproget ordbog	s. 371
PÅR NILSSON: En kognitiv semantisk analys av SAOB:s formler för betydelseförändring	s. 383
CARINA NILSTUN: NAOB: På vei mot en nasjonalordbok for bokmålet: Nyordprosjektet ”2014 nyord i 2014”	s. 395
SANNI NIMB & BOLETTE SANDFORD PEDERSEN: Fra begrebsordbog til sprogteknologisk ressource: verber, semantiske roller og rammer – et pilotstudie	s. 405
BOLETTE SANDFORD PEDERSEN, SANNI NIMB, ANNA BRAASCH & SUSSI OLSEN: Betydningsinventarer – i ordbøger og i løbende tekst	s. 417
EVA THELIN: Rietz och riksspråket	s. 431
LENA WENNER: Vilka dialektord kvalar in i en standardspråklig ordlista?	s. 443
KONFERENCEPROGRAM	s. 455

Forord

Den 13. Nordiske Konference i Leksikografi blev afholdt i dagene 20.-22. maj 2015 på Københavns Universitet. Arrangørerne var Nordisk Forening for Leksikografi og Ømålsordbogen, Nordisk Forskningsinstitut, Københavns Universitet. Den lokale arrangementskomite bestod af Asgerd Gudiksen, Henrik Hovmark, Liisa Deth Theilgaard og Lena Wienecke Andersen. Komiteen har undervejs fået nyttige råd fra Rikke Hauge, og under selve konferencen var studentermedhjælperne Anne Fröling Hansen, Bente Kristensen og Albert Worsøe en god støtte ved afviklingen af det praktiske arrangement.

Konferencen samlede 88 deltagere fra hele Norden. Der var repræsentanter fra mange forskellige institutioner. Og det afspejlede sig tydeligt i konferenceprogrammet, der indeholdt meget varierede bidrag. Konferencen havde som noget nyt for de nordiske leksikografikonferencer et tema: *Den digitale ordbogsbruger*, men også foredrag uden for temaet var velkomne. Der blev holdt 46 præsentationer: 4 plenarforelæsninger, 38 sektionsforedrag og 4 posterpræsentationer. Præsentationerne delte sig mellem 27 tematiske bidrag og 19 ikke-tematiske. Under konferencen var der også arrangeret en udstilling af ordbøger og andre leksikografiske værker fra nordiske forlag og institutioner.

Konferencens sociale program indledtes med en uformel sammenkomst den 19. maj på Nordisk Forskningsinstitut. Under sammenkomsten var der arrangeret rundvisning på *Ordbog over det norrøne Prosasprog* og på *Ømålsordbogen*. Den følgende eftermiddag var der rundvisning på Nordisk Film og besøg på udstillingsstedet Cisternerne i Søndermarken på Frederiksberg, hvor der var installationer af den svenske kunstner Ingvar Cronhammar. Konferencemiddagen blev holdt på Restaurant Bellahøj, der oprindeligt blev bygget som en lystgård i 1792. Under konferencen blev der også afholdt generalforsamling i Nordisk Forening for Leksikografi.

Nordiske Studier i Leksikografi 13 indeholder 35 artikler udarbejdet på grundlag af de afholdte foredrag. Oddrun Grønvik, en af plenarforelæserne, har dog fået mulighed for at bringe en artikel der er udvidet og omarbejdet i forhold til det oprindelige foredrag. Bogens indhold falder i tre dele: Først to artikler skrevet af plenarforelæserne, dernæst 17 artikler inden for temaet *Den*

digitale ordbogsbruger og til sidst 16 artikler med andre emner. Artiklerne er med en enkelt undtagelse på et af de skandinaviske sprog og forsynet med et engelsk resume. Artiklerne inden for temaet indeholder både brugerundersøgelser, overvejelser om hvordan man imødekommer brugernes behov bedst muligt og løsninger valgt af ordbøger af vidt forskellige typer. Den ikke-tematiske del rummer artikler om fx sprogteknologisk baserede ordforrådsstudier, sproglige domæner, ordbogshistorie, leksikografisk teori og leksikologiske emner. Som helhed afspejler bogens artikler den faglige spændvidde som kendetegner det nordiske leksikografiske miljø, og giver samtidig et godt indblik i den aktuelle diskussion om brugernes rolle i leksikografien i en digital tidsalder.

Konferencerapporten er redigeret af Asgerd Gudiksen og Henrik Hovmark. Henrik Hovmark har også stået for satsen med teknisk bistand fra Liisa Deth Theilgaard og Bo Nissen Knudsen, som vi begge takker varmt.

Afslutningsvis skylder vi også alle de anonyme fagfællebedømmere en stor tak. De har med stor beredvillighed påtaget sig den ikke helt lille opgave at vurdere og kommentere en videnskabelig artikel. Vi takker også forfatterne for godt samarbejde. Sidst, men ikke mindst vil vi takke Nordisk Forskningsinstitut og Nordplus Nordiske Sprog, som har ydet uundværlig finansiel støtte til både konferencen og udarbejdelsen af konference-rapporten.

København, juni 2016

Asgerd Gudiksen og Henrik Hovmark

PLENARFOREDRAG

Ord og IT

Esben Alfort

With so many new digital language tools being developed these years, it is easy to fear for the immediate future of traditional lexicography. However, no digital tool can be perfected without lexicographic knowledge being supplied by someone who really understands the way of words. In fact, the need for lexicographic skills has never been greater.

1. Et nyt syn på tingene

Da jeg sidste år var blevet inviteret til at sidde i et panel som skulle diskutere ordbogens fremtid, gjorde jeg det meget klart at jeg som repræsentant for den sprogteknologiske virksomhed Ankiro føler mig overbevist om at de traditionelle ordbøger også har en rolle at spille fremover. Ganske vist dukker der hele tiden nye smarte digitale værktøjer op som giver unikke muligheder for at udforske sprog og finde frem til den helt rigtige betydningsnuance, men det ændrer ikke ved det faktum at det traditionelle ordbogsformat, hvad enten det er på papir eller som elektronisk dokument, har sine styrker og klare fordele, og derfor vil der utvivlsomt også fremover være efterspørgsel efter dem. Alligevel er der ingen tvivl om at brugen af især papirordbøger er faldet mærkbart med den øgede adgang til mobile elektroniske alternativer. Derfor sker det også let at man som leksikograf bukkes under for en ulmende frygt for at de digitale værktøjer helt vil overtage markedet og tage livet af ordbogsbranchen. Hvis man udelukkende ser de elektroniske sprogværktøjer som en trussel, går man imidlertid glip af et stort potentiale for gensidigt givende samarbejder, for der er stor brug for folk med leksikografisk og anden sproglig viden og kunnen inden for den sprogteknologiske branche. Computere er notorisk uintelligente, så uden denne viden og indsigt bliver værktøjerne aldrig tilfredsstillende.

Selv er jeg uddannet teoretisk lingvist fra Københavns Universitet og havde egentlig forestillet mig en helt anden karriere da jeg ved et tilfælde havnede på Ankiro for 7-8 år siden, hvor der var brug for en svenskkyndig til at lave det der skulle blive vores svenske digitale ordbog. Siden lavede jeg

en finsk ordbog, og med tiden fik jeg ansvar for en række forskellige projekter, bl.a. udtræk af forskellige slags information ved hjælp af natursprogs-parsing. I dag er jeg medejer og leder af forskningsafdelingen, som opret- holder nære kontakter til universitetsverdenen. Jeg blev meget hurtigt fænget af atmosfæren i Ankiro, både fordi firmaet har et stærkt fokus på fordybelse og forskning, og fordi der opstår en utroligt kreativ atmosfære når man sætter netop sprogfolk og programmører sammen. Det er to meget forskel- lige mennesketyper med et vidt forskelligt syn på næsten alt, men ikke desto mindre med en fælles grundlæggende fascination af *systemer* og en udviklet strukturforståelse. Forskellene betyder i praksis at man er tvunget til at ny- tænke og redefinere alting før man bliver i stand til at formidle sine tanker på tværs af afdelingerne. Det lyder måske besværligt, men resultatet er at begge parter forstår fænomenerne langt bedre, hvad enten det drejer sig om sprog eller IT.

Ved at fortælle lidt om hvad vi laver i Ankiro og hvad det er for nogle sproglige udfordringer vi dagligt brydes med, håber jeg at det vil lykkes mig at formidle det store behov der er for sprogfolk som os – inkl. leksikografer – i den digitale verden.

Ankiro lever af at lave sprogteknologi – et begreb der dækker over en lang række forskellige løsninger og produkter, alt efter hvad vores kunder har brug for. Fælles for alle vores løsninger og produkter er at de involverer kombinationen af IT og sprog, hvilket nærmest pr. definition kan være tem- melig udfordrende, og fælles for os der arbejder i Ankiro, er derfor også at vi elsker sproglige udfordringer. Opgaverne skal faktisk helst være på grænsen til det umulige. Vi opererer primært på dansk, svensk, bokmål, nynorsk og finsk, samt i skrivende stund engelsk, tysk, hollandsk og fransk for de kun- der der har brug for det, men Norden har altid været vores kerneområde.

Det hele startede i 1999 med intelligent søgning. Idealet har fra starten været at vores søgemaskiner skal arbejde på brugernes præmisser og finde det de leder efter og ikke bare det de skriver, og det er kun muligt ved at tilknytte en række sproressourcer i form af synonymordbøger, ontologier og forskellige slags parsere og regelværktøjer. Der er jo ingen garanti for at brugerne formulerer sig på samme måde som forfatteren af de dokumenter de leder efter, eller for den sags skyld at de overhovedet tilstræber at efter- ligne sproget i det søgte idealdokument.

Når man arbejder med sprog, må man desuden altid være forberedt på at ordene har en ekstra overraskelse i baghånden. Hvad er f.eks. forskellen på *lakridsstang* og *stanglakrids*? I hvilken udstrækning er man interesseret i det ene hvis man søger efter det andet? Den slags materiale-og-form-ord (i Ankiro kaldet ”stanglakridsord” – andre eksempler er *sandstrand* og *kursus-uge*) er bare en af mange sproglige finurligheder der må håndteres med en særlig regel specialdesignet til netop det formål.

Ret hurtigt fandt man i Ankiro på at bruge de ressourcer man havde udviklet til søgning til også at lave nogle store jobportaler. Hver nat gennemsøger vi det danske internet for alt hvad der ligner et jobopslag, og derefter gør vi det muligt for jobsøgende at søge i databasen med semantiske hjælpemidler. Man kan f.eks. søge på *leksikograf Amager* og få foreslået et job som ordbogsredaktør i København, fordi systemet ved at en leksikograf næsten er det samme som en ordbogsredaktør og at København ligger tæt på Amager. Projektet blev en stor succes og udgør i dag Danmarks største jobportal.

En anden ting vi bruger vores sprogressourcer til, er informationsudtræk af forskellig art. Vi analyserer store tekstkorpora automatisk og udtrækker bestemte oplysninger ved hjælp af møjsommeligt udformede abstrakte sproglige regler. Igen er det kun muligt fordi vi kan trække på de ordbøger og ontologier m.m. som vi har udviklet gennem årene og fortsat løbende forbedrer. Man kan nemlig ikke gøre reglerne tilstrækkelig abstrakte uden at tilføje viden om synonymi, taksonomi og syntaktiske variationsmuligheder.

Vi ligger inde med de sidste mange års jobopslag fra hele landet, og det er selvfølgelig en guldgrube af viden om hvad man forventes at kunne hvis man søger bestemte stillinger. Jeg er derfor begyndt at anvende vores informationsudtræksmetoder på disse jobdata og løbende udtrække kompetencekrav fra jobopslagene for fagbevægelsen så de jobsøgende kan blive vejledt endnu bedre. Arbejdet bliver dog aldrig færdigt, for det er et evigt kapløb med forfatterens sproglige kreativitet og variationslyst. Man når lige at lære systemet at fortolke sætninger som *desuden skal du være interesseret i og helst de sidste mindst fem år have arbejdet intenst med rørtrådssvejsning*, før man mødes af formuleringen *du har en ivrig rørtrådssvejsner i maven...*

2. Udfordrende ord

Også søgemaskiner er som sagt afhængige af sproglig viden hvis de skal leve op til idealet om at fungere på brugernes præmisser. Søgning er jo kommunikation mellem en bruger og et system, og derfor kommer der også alle mulige sproglige forhold ind som komplicerer processen og som man er nødt til at finde løsninger på. For det første indeholder 10-20 % af alle søgninger mindst én stavfejl, afhængig af domæne og målgruppe, og de kan sommetider være meget alvorlige. Hvad menes der f.eks. med ordet *tredbrøn*? Man kan måske få en idé når man ser den fulde søgestreng *trådløs tredbrøn*, især når man ved at søgningen blev udført på hjemmesiden for et firma der leverer bredbåndsforbindelser, men det er ikke nemt for en computer at afgøre.

For det andet udtrykker folk sig som nævnt ikke nødvendigvis på samme måde som den virksomhed eller institution der ejer det site de søger på. Meget ofte vælger de et andet ord end det der bruges i det dokument de leder efter, og tit bruger de desuden andre morfologiske former og syntaktiske konstruktioner. Her følger to eksempler.

- | | |
|-----------------------|----------------------|
| 1. enddentetstyveri | 15. identitetstyveri |
| 2. id tyv | 16. identets tyveri |
| 3. id tyveeri | 17. identetsstyveri |
| 4. id tyveri | 18. identetstyveri |
| 5. idenditets tyveri | 19. identetstyveri |
| 6. idenditetstyver | 20. identetstyveri |
| 7. idenditetstyveri | 21. identstyveri |
| 8. idenditetstyveri | 22. identstyveri' |
| 9. ideniitets tyveri | 23. ididentetstyveri |
| 10. idensitets tyveri | 24. IDtyveri |
| 11. idensitetstyveri | 25. idtyveri |
| 12. identetstyveri | 26. indentet tyveri |
| 13. Identetstyveri | 27. indentetstyveri |
| 14. identet tyveri | 28. indentetstyver |

- | | |
|-------------------------|--------------------------|
| 1. rotte | 10. rottejagt |
| 2. rotter | 11. rotteudryddelse |
| 3. rotter på loftet | 12. rotte bekæmpelse |
| 4. skadedyr rotter | 13. rotte bekæmpelse |
| 5. rotteplage | 14. rottebekæmpelse |
| 6. anmeldelse af rotter | 15. rottebekæmpelse |
| 7. rottefanger | 16. rottebekæmpelse |
| 8. rottefænger | 17. rottebekæmpelse |
| 9. rottemand | 18. bekæmpelse af rotter |

Det første materiale stammer fra en søgeløg knyttet til en stor dansk virksomheds intranet, hvor jeg fandt 28 forskellige måder at referere til identitetstyveri på. Ikke alle er stavefejl; for at fortolke disse udtryk skal systemet både bruge information om synonymmer, forkortelser og afledninger, foruden en kraftig intelligent stavekontrol.

Det andet eksempelmateriale er hentet fra en kommunal søgeløg og består af 18 forskellige måder at søge om hjælp til rottebekæmpelse på.

Der er en hel del sproglige udfordringer bare i dette lille materiale – stavefejl, morfologisk variation (ental vs. flertal), syntaktisk variation (*rottebekæmpelse* vs. *bekæmpelse af rotter*), synonymvalg (*bekæmpelse* vs. *udryddelse*), og ren og skær forskel i tilgang (*rotteplage* vs. *rottefænger* vs. *rottebekæmpelse*). De personer der har foretaget alle disse søgninger har formentlig alle sammen haft omtrent samme intention – de vil finde information om rottebekæmpelse, og sandsynligvis også anvende denne information til at tilkalde hjælp, men de går til opgaven på helt forskellige vis.

Stavekontrollen er bygget op omkring en kombination af tastetekniske og fonetiske regler. De tastetekniske regler tager højde for at man kan ramme forkert, så bogstaver der ligger tæt ved hinanden, eller f.eks. i samme finger i modsat hånd, kan blive forbyttet, mens de fonetiske regler afspejler det enkelte sprogs lydssystem og specificerer de mest oplagte muligheder for ortografisk forveksling af ens eller næsten ens lyde. Disse regler er det første vi laver, når vi tager hul på et nyt sprog.

Selv med sådan et kraftfuldt stavekontrolapparat er man dog aldrig helt sikker, for konteksten kan også spille ind i fortolkningen, som vi så med

bredbåndseksemplet. Hvis man som menneske ser søgestrengen *kartofler carutter*, så er man ikke i tvivl om at der skulle have stået *karotter*, men hvis søgestrengen havde været *cigarer carutter*, så ville man have tolket det samme ord som *cerutter*.

Hvis man vil være i stand til at tage højde for konteksten i den slags situationer, er man nødt til at modellere det relevante emneområde. Til det formål opbygger man en ontologi, dvs. en formel struktur af begreber bundet sammen af forskellige relationer, f.eks. *kartofler* og *karotter* som underbegreber til én knude og *cigarer* og *cerutter* under en anden. Ofte vil kun den ene betydning være relevant for en given kunde, fordi den anden falder uden for domænet. Så kan konteksten være underforstået på det sproglige plan fordi den er knyttet til sitet som helhed.

Den mest klassiske måde at opbygge ontologier på er som en taksonomi af over- og underbegreber; når en zebra er et pattedyr og et pattedyr er et dyr, så er det måske oplagt at en der søger efter information om dyr også potentielt vil være interesseret i information om zebraer. Helt så simpelt er det desværre ikke, for information kan være generel eller specifik, og i nogle tilfælde er det decideret problematisk at inddrage underbegreber i søgemaskineriet.

Hvis man søger efter dokumenter der handler om *kæledyr*, er det ikke utænkeligt at man i mange tilfælde også kunne være interesseret i at finde information om katte. Man kunne derfor lægge *katte* ind i ontologien som underbegreb til *kæledyr*. Man kunne også lægge en række andre dyrebetegnelser som *hund*, *marsvin* og *rotte* ind, men man skal passe meget på, for det kan hurtigt komme til at skade mere end det gavner. Hvis man søger efter information om kæledyr på Københavns Kommunes hjemmeside, er man f.eks. næppe interesseret i at høre om rotter, for med kommunens briller er en rotte altid et skadedyr, og hvis man søger efter information om kæledyr, er man mindre interesseret i at få at vide hvordan man bedst slår dem ihjel...

Det er således kontekst- og domæneafhængigt hvad der skal inkluderes som underbegreber. Hvad der er farligt at have med, skifter fra domæne til domæne, så man er nødt til at opbygge forskellige ontologier til forskellige domæner, og i det hele taget være meget forsigtig og altid tænke på hvad informationen skal bruges til i et konkret tilfælde.

I Ankiro har vi derfor en række *kundeordbøger* designet til at passe den enkelte kundes behov. De er baseret på nogle mere generelle *domæne-*

ordbøger, som igen er dannet på baggrund af *almensproglige ordbøger*, én for hvert sprog, og tilsvarende med ontologier. De almene ordbøger er 100 % manuelt gennemarbejdede i løbet af de sidste 15 år (for den danske ordbogs vedkommende, kortere for de andre sprog). Vi kunne også have brugt de 15 år på at udvikle et automatisk statistikbaseret system, men så ville vi have stået med et langt mindre solidt og pålideligt materiale i dag. Det har derfor altid været vores filosofi at alt skal være tjekket af sprogfolk, i hvert fald i de almensproglige ordbøger som bruges på tværs af de enkelte projekter, og hvis indhold man derfor skal kunne stole på.

Eksemplerne ovenfor viser i øvrigt at man ikke bare kan importere ontologier og tesauri fra eksterne ressourcer, for så får man alt for mange relationer. Man kan lade sig inspirere af dem, men ikke importere dem ukritisk. Da den engelske ordbog i sin tid blev etableret, købte man sig til en synonymordbog, og så måtte man ellers bruge enorme ressourcer på at slette semantiske relationer, fordi det simpelthen ikke er holdbart at man får resultater om alt muligt som på en eller anden måde er relateret til det man har søgt på, eller synonymt med det i en eller anden begrænset kontekst. Der skal meget strengere kriterier til, baseret på den konkrete anvendelse materialet er tiltænkt. Det kan jo ikke nytte noget at man får tekster om kartofler når man søger efter information om kæledyr, bare fordi *fritter* både kan være kæledyr og franske kartofler. Så man må være meget tilbageholdende og bruge sin sproglige fornemmelse.

Når stavekontrol og semantiske relationer er på plads, kommer næste trin i forståelsesprocessen, nemlig det vi kalder *konstruktionsgenkendelse*. Det er semantiske relationer på konstruktionsniveau, eller ”syntaktisk synonymi”. Visse konstruktioner er synonyme eller beslægtede, f.eks. *bekæmpelse af rotter* og *rottebekæmpelse*. Selvom dette er en meget oplagt synonymi-relation, kan en computer ikke uden videre se det; man er simpelthen nødt til at lære systemet at visse typer af verbalsubstantiv fulgt af visse præpositionsforbindelser er synonyme med det omvendte kompositum. Den slags følger i vid udstrækning faste regler, men sprog er jo altid fulde af undtagelser, så man er alligevel nødt til at lægge en hel del sproglig information ind for ikke at få for meget støj. Hvordan skal systemet f.eks. vide at det hedder *forbud mod* og ikke *forbud af*, og hvad indebærer det for et ord som *havbeskyttelse* at samme korpus kan indeholde ordet både i betydningen *beskyttelse AF havet* og *beskyttelse MOD havet*? Den slags spørgsmål vil det altid kræve

ægte sprogfolk at besvare. I sidste ende gør man således klogt i at lade de automatiserede metoder være vejledende snarere end bestemmende i sig selv.

Det er tydeligt at et intelligent søgesystem har brug for en hel del sproglig information af alle mulige slags for at kunne håndtere selv de simpleste søgninger. Faktisk er der behov for endnu flere slags information fra den sproglige verden end man måske umiddelbart skulle tro. Ud over de nævnte fænomener er der nemlig nogle ret uventede som man også skal tage stilling til. Hvis man f.eks. går ind på sitet *visitlondon.com* og søger efter information om *Arsenal*, men kommer til at skrive *Arsenol*, svarer den med spørgsmålet *Did you mean arsehol?* Hvis man på Google søger på *perlersvin* og glemmer et mellemrum, så bliver man tilsvarende spurgt om man i virkeligheden mente *perkersvin*. Den slags forslag er ikke altid lige heldige, især ikke hvis det site man befinder sig på, tilhører en offentlig myndighed, for så kan det opfattes som en blåstempling af det foreslåede udtryk. Ankiros søgemaskiner ved godt at der er visse udtryk man skal holde sig fra at smide i hovedet på brugerne. De er ganske enkelt markeret som potentielt anstødelige i vores ordbøger.

Der skal altså også lægges stiloplysninger i ordbøgerne, og helst også information om genre og domæne, hvis man vil vide hvad der er passende i en given sammenhæng. Det forudsætter selvfølgelig at der er nogen der lægger informationen ind, og her kommer den leksikografiske viden endnu engang på banen. Der skal altid lidt ekstra til for at få ord og IT til at fungere sammen.

3. Intention

Man kan spørge sig selv hvad det er der får folk til at udtrykke sig forskelligt. Hvorfor udtrykker de sig som de gør, og hvad er egentlig forskellen på de forskellige udtryk? Ligger der en forskel i intention bag en forskel i udtryk? Det er vigtigt at vi stiller disse teoretiske spørgsmål, for hvis vi kunne forstå hvad der får folk til at udtrykke sig som de gør, så kunne vi måske også vende processen om og forstå hvad de har behov for ud fra den måde de udtrykker sig på. Men det er aldrig helt let.

Hvad er det for eksempel der får visse brugere til at skrive *har glemmt min pinkode*, mens andre skriver *ny pinkode* og atter andre *lav ny pinkode*? Det er tre forskellige former for syntaks der anvendes, men ligger der så også en tilsvarende forskel i intention bag dette valg? Når det første er en beskrivelse af problemet, mens de andre to er forsøg på at finde en løsning, betyder det så at brugerne konceptualiserer problemet forskelligt, og måske også har lidt forskellige behov, eller interagerer de bare forskelligt med systemet? Og hvordan er det lige med deiksis – ligger der noget bag perspektivvalget i udtrykkene *betale MIN regning*, *betal DIN regning* *HER* og *kan man betale SIN regning*?

Det var sådan nogle spørgsmål der fik mig til at kaste mig ud i det forskningsprojekt der mandede ud i en ph.d.-afhandling, som var finansieret af Ankiro og CBS i fællesskab, hvor jeg analyserede nogle af vores kundesites søgelogs for at se hvordan folk udtrykker sig og i første omgang udvikle et framework og en notationsform som gør det muligt at diskutere variationen på et videnskabeligt plan.

Noget af det der gør det så vanskeligt at fortolke søgestreng, er at folk er notorisk kortfattede når de formulerer sig i et søgefelt. Typisk består en søgestreng af 1-2 ord, og det giver jo ikke systemet (og analytikeren) ret meget at arbejde med, hverken hvad angår sproglig substans eller kontekst. Alligevel forventer brugerne at få et fornuftigt svar på deres forespørgsel. Det er egentlig ret meget forlangt, for man går jo ikke bare ind i en butik og siger ”Hurtig pc” og forventer at få fornuftig betjening på basis af det cue, og man ringer heller ikke til et supportcenter og siger ”Langsom pc”. Ikke desto mindre er det sådan folk udtrykker sig når de søger, og de forventer at det virker, og det skal være inden for få millisekunder.

De to nævnte udtryk, *hurtig pc* og *langsom pc* er i øvrigt ret interessante, for selvom de er helt identisk opbygget, så dækker de over vidt forskellige intentioner, hvilket da også blev afspejlet i mit valg af eksemplificerende situationer: mens *hurtig pc* er en produktsøgning, er *langsom pc* en support-søgning hvor man søger en løsning på et problem. Hele forskellen ligger i det leksikalske valg af adjektiv. Det er måske ikke så underligt at et adjektiv med den modsatte betydning kan ændre betydningen af et udtryk drastisk, men hvordan skal computeren vide at der pludselig er tale om en helt anden type søgning og ikke bare et andet emne? Svaret er selvfølgelig, at nogen er nødt til at fortælle den det. Nogen med forstand på ord og sprog.

4. Reference, modalitet og aspekt i søgning

Hvis man søger på *leksikograf Amager* i en af Ankiros jobportaler, er systemet som nævnt i stand til at fortolke *leksikograf* som en stilling og *Amager* som et geografisk sted og kan finde relaterede jobs i nærheden. Hvad systemet ikke kan vide, er om brugeren mente at hun *vil være* leksikograf på Amager, eller at hun *er* leksikograf fra Amager og gerne vil have et relevant job i nærheden. Det kan jo være afgørende, især fordi vi går ind og analyserer på hvilke kompetencer der står i jobopslagene at man skal have og kan anbefale kurser hvis der er noget vigtigt der mangler på cv'et.

Man kan i den forbindelse spørge sig selv hvad søgestreng overhovedet refererer til. Kan man gøre sig a priori antagelser om hvad folk vælger at referere til? Den traditionelle opfattelse er at søgestreng refererer til *emner*, eller ”topics”, men så enkelt er det faktisk ikke. I en vidensbase er det ganske rigtigt oplagt at emnesøgninger er hyppige, men i en job- eller boligportal møder man dem kun sjældent, og i en webshop er de så godt som ikke-eksisterende; her det andre typer af søgninger der dominerer. Brugerne har med andre ord et bredt spektrum af referencemuligheder til rådighed.

Grundlæggende viser det sig at man kan referere enten til det jeg kalder Stadiet 1 – den problematiske situation – eller Stadiet 2 – den ønskede situation hvor problemet er løst. Et eksempel er jobsøgningssituationen fra før, hvor man kan beskrive den nuværende tilstand (hvad man kan og hvor man bor) eller den ønskede tilstand (det job man er interesseret i og det sted man ønsker at arbejde). Et andet eksempel er supportsituationen, hvor man kan beskrive problemet eller den løsning man forestiller sig kunne findes og som man ønsker information om. Søgstrengen *langsom pc* er en Stadiet 1-søgning der beskriver problemet; en tilsvarende Stadiet 2-søgning kunne f.eks. være *hurtigere pc*.

Med lingvistiske briller kan man sige at stadiet 1 er en form for modal distinktion mellem det deklarative som gælder før søgningen og det optative som man ønsker skal gælde i fremtiden. Man kan så referere til en af disse to situationer. Desværre kan et søgesystem aldrig vide om brugeren refererer til stadiet 1 eller stadiet 2, fordi referencen udføres på nøjagtig samme måde. En løsning kunne selvfølgelig være at spørge brugeren direkte, men folk er som regel ikke ret glade for at skulle svare på spørgsmål når de søger. De vil bare have et resultat. Og det skal være nu.

Det viser sig at der er flere interessante egenskaber ved den måde der refereres på. Foruden den *modale* forskel mellem den eksisterende problematiske situation og den ønskede eller potentielle løsning er der også *aspektuelle* forskelle som afhænger af hvordan brugeren anlægger sit perspektiv på de to situationer. Det er noget der er meget fremtrædende i forbindelse med jobsøgning, hvorfor jeg vil tage mine eksempler fra det domæne.

Mit eksempel kommer fra WorkInDenmark, som er et offentligt site for udlændinge der søger job i Danmark. En analyse af søgeloggen viser med det samme at der er store forskelle på hvordan man griber en jobsøgning an. Kun i godt halvdelen af tilfældene refererer brugerne til det job de søger, og selv blandt dem er der stor variation. Nogle beskriver stillingen, det sted de vil arbejde eller det de vil arbejde med. Engelsksprogede eksempler er *driver* (stillingsbetegnelse), *cleaning* (arbejdsopgave), *farm* (arbejdsrammer) og *pig* (materiale m.m.). Det er alt sammen Stadio 2-søgninger med et "neutralt" aspekt, men af forskellige undertyper. Der findes imidlertid også en anden slags Stadio-2 søgninger som har hvad man kunne kalde et *prospektivt* aspekt; det er folk der beskriver det søgte job som set fra deres egen synsvinkel, som noget der udfylder et eksisterende behov. Eksempler på dette er *student job*, *seasonal work* og *copenhagen*. Sådan nogle søgninger fortæller os ofte mere om den søgende selv end om jobbet, for hvis man søger job i København, er der stor sandsynlighed for at det er fordi man bor (eller har tænkt sig at bo) i eller nær København og derfor søger noget i nærheden, og hvis man søger studiejob, er det nok tilsvarende fordi man er studerende.

Omvendt er der også nogle der beskriver sig selv som en potentiel arbejdsgiver ville se dem, som værende personer med forskellige kompetencer. Det er Stadio 1-søgninger med et *retrospektivt* aspekt. Eksempler fra WorkInDenmark er *no experience*, *biology*, *java* og *spanish speaking*. Man kan selvfølgelig argumentere for at disse søgninger er eksempler på at folk leder efter jobopslag der indeholder disse strenge, men den mulighed er altid til stede uanset hvordan søgningerne er udformet og er derfor helt uinteressant fra et analysesynspunkt. Hvis der er en anden mulighed, bør man som loganalytiker næsten altid vælge den, fordi det tilfører mere information om den søgende. Alternativet er ganske enkelt at konstatere at loganalyse er umuligt, og det bliver vi jo ikke klogere af.

Den sidste hovedtype af jobsøgningsreferencer repræsenteres af dem der beskriver sig selv uden nogen som helst reference til arbejdsgiveren. Det er

folk der ganske enkelt refererer til sig selv som *spanish*, *young* eller *unemployed*. Man kunne måske tro at denne type var sjælden, men faktisk er det nogle af de hyppigste søgninger på hele sitet. Tallene er dog lidt misvisende, for en betegnelse som *spanish* vil naturligvis oftere være at betragte som en sproglig kompetence end en nationalitetsbeskrivelse. De fem hyppigste søgninger på sitet er *spanish*, *engineer*, *french*, *russian*, *farm* og *english*, og det er tydeligt at de nævnte nationaliteter netop er nogle der tit vil optræde som sprogkompetencer, men lidt længere nede på listen finder man mange der ikke umiddelbart kan tolkes på den måde, f.eks. *nepal*, *mexican* og selvfølgelig *young*.

Der er tydeligvis stor forskel på hvordan man går til selv en så klart defineret opgave som at søge job. Faktisk opererer jeg i min afhandling med 22 undertyper af jobsøgning, så man skal først og fremmest være forberedt på stor diversitet i brugen af enhver søgemaskine. På den baggrund siger det sig selv at en mindre klart defineret søgning som f.eks. søgeboksen på et site eller et intranet har et meget komplekst og indirekte forhold til den intention der ligger bag en given søgning.

5. Konklusion

Tilbage er kun at konstatere at ord og IT er en vanskelig kombination. Faktum er at der mere end nogensinde er brug for sproglig ekspertise og digitale sproglige ressourcer ude hos os der laver de digitale værktøjer, og det får man kun med sprogfolk som os. Jeg vil derfor benytte lejligheden til at opfordre til nye frugtbare samarbejder. De digitale værktøjer er ikke en trussel mod leksikografien; de repræsenterer nye muligheder for os alle sammen. Udfordringen er så at identificere de nye behov, og de dækker i vid udstrækning stadig over traditionel leksikografisk viden og kunnen.

Litteratur

Alfort, Esben (2013): *The expression of a need – Understanding search*. (PhD thesis). Frederiksberg: Copenhagen Business School, Department of International Business Communication.

Nordiske Studier i Leksikografi 13 (2016): 13-25

Esben Alfort
leder af Afdelingen for Forskning og Innovation, ph.d.
Ankiro
ealfort@ankiro.dk

Vitskaplegheit og samfunnsrelevans for store ordboksverk

Oddrun Grønvik

Norsk Ordbok, a 12 volume scholarly dictionary of the Nynorsk written language and all Norwegian dialects, was completed in 2015 and launched in March 2016. The publishing history of Norsk Ordbok lasted for 66 years, with 4 volumes published in the first 52 years and 8 in the following 13, after the dictionary was redefined as a project and moved to a digital platform. With this publishing history as its basis, this article explores the development of scholarly lexicography in terms of scientific theory and technology underlying macrostructure, microstructure and the genre itself. Finally relevance, legitimacy and societal and scientific worth of scholarly lexicography are briefly summed up.

1. Innleiing

Onsdag 9. mars 2016 vart det feira i stor stil i Oslo at *Norsk Ordbok. Ordbok over det norske folkemålet og det nynorske skriftmålet* (NO) var ferdigstilt frå a til å, i 12 band. Arbeidet tok 87 år frå materialinnsamlinga byrja og til siste band kom frå trykkeriet. NO-redaksjonen fullførte manuskriptet i 2015 og fekk siste band på prent i 2016.

NO er det einaste av det store vitskaplege ordboksverka i Europa som handsamar skrift og tale som likeverdige kjelder for ordtilfanget. Heilskapen er på opp mot 67 millionar teikn, som svarer til 225 romanar av vanleg storleik. Meir enn 300 000 ord har fått eigne ordartiklar, om lag halvparten for første gong. Meir enn femti morsmålsleksikografar har vore med på ferda sidan 1930, mange mellom dei fagleg mest framståande norskfilologane i si tid (sjå kortbiografiar i Karlsen et al. 2016:170 ff.). I redigeringsperioden 1946-2016 er nynorsk blitt utvikla til å bli eit fullført skriftmål og ein konstituerande del av norsk skriftverd. For all framtidig norsk leksikografi bør NO vera eit uomgjengeleg referanseverk.

Det er heva over tvil at overgang til redigering til digital plattform har vore avgjerande for fullføringa av NO. I denne artikkelen vil eg gå inn på

kva denne overgangen inneber, både vitskapleg og praktisk. I samband med utforskinga av dette emnet vart det naudsynt å sjå på dei lange linjene i utviklinga av ordboksformatet som verktøy for beskriving av sambandet mellom ordtilfang og tydingsregister.

Eit tilknytt emne er spørsmålet om leksikografiens vitskaplege status. I samband med at leksikografi er lagt ned som disiplin ved Universitetet i Oslo, er det blitt hevda at empirisk leksikografi ikkje er eit forskingsfelt for lingvistikken (sjå nærare omtale av denne diskusjonen i Vikør 2016:122). Mot dette vil eg seia at leksikonet er like mykje av eit forskingsfelt som andre sider ved språket. Dei store vitskaplege ordboksverka er å rekna som grunnforskning, av di dei utgjer ei først gjennomarbeiding og fortolking av store og komplekse materialsamlingar.

2. Ordboksverk og vitskaplegheit

Vitskaplege ordbøker har same modell som alt anna vitskapleg arbeid med empirisk grunnlag. Med utgangspunkt i ein teori, og eit gjeve og definert tilfang, isolerer ein det fenomenet som skal granskast, gjer greie for metode, og byggjer opp ei systematisk framstilling av dei ulike instansane av fenomenet. I ordbøker, er fenomenet ordtilfanget (i ein eller fleire språklege varietetar), instansane er oppslagsorda og framstillinga er ordartiklane. Metode skal vera gjort greie for i retningslinjer for redigering, føreord osv.

Fleire av dei store vitskaplege ordboksverka som skildrar moderne europeiske språk gjennom morsmålsbriller, ligg ved institusjonar utanfor universiteta. Grunnane er samansette, men nokre fell ein i auga: a) Ordbokstiltaka kom i gang før morsmåla var blitt akademiske undervisnings- og forskingsemne i sine land.¹ b) Ordboksverka krev oppbygging av arkiv og samlingar som universitetsstrukturen normalt ikkje har rom for.² c) Dei store ordboksverka som kom i gang før 1940, mangla realistiske gjennomføringsplanar og sleit i konstant pengenaud og mannskapsmangel. Dei vart fleir-

¹ OED vart vedteke som prosjekt i 1857 og kom i gang for alvor frå 1879, medan The School of English ved Universitetet i Oxford kom i gang frå 1893. Jf. Murray 2001:71; Palmer 1965.

² Ivaretaking av grunnlagsdata er ei pågåande sak i Noreg pga Forskingsrådets krav om lagring av forskingsdata i tilgjengeleg format i minst 10 år, om om plan for langtidslagring av utvalde data (Noregs Forskingsråd 2014: 5).

generasjonsverk, men sidan vitskapleg arbeid krev konsistent gjennomføring, vart dei etterkvart rekna som talsmenn for forelda vitskapleg metode. Samstundes kunne redaksjonar motsetta seg endring, fordi dei var redde for kvalitetssvikt. (Vikør 1999:397). Dette har nok gjort sitt til at universiteta har vore skeptiske til å hysa ordboksverk.³

I dag forstår alle at ordboksverk er produkt av si tid. Fleirgenerasjonsmodellen er eit tilbaketog stadium for alle slags forskingsprosjekt. Nye ordboksprosjekt vil bli planlagde med tanke på rask fullføring og stram kontroll. I dei siste tjue åra har vi fått fleire døme, ikkje minst i Norden, på at kvalitet og effektivitet kan sameinast. Eitt slikt døme er *NO*.

Har dei store ordboksverka ein plass i forskingsverda? Dersom ein meiner at leksikon er jamgodt med fonologi, morfologi og syntaks som forskingsemne, og fortener vitskapleg handsaming på same nivå, er svaret ja. Samlingsoppbygging og redigering krev spesialisert vitskapleg kompetanse, og denne kompetansen bør også koma dei tradisjonelle verkeområda ved universiteta til gode. Korleis dette skal organiserast er eit spørsmål som truleg krev eiga løysing for kvart ordboksprosjekt.

Dei vitskaplege ordboksverka har ein eigenskap som vitskaplege bidrag innanfor lingvistikkens deldisiplinar manglar; dei gjev ein syntese av deldisiplinane, i einskildartiklar og i høve til einskildord. Som heilskap gjer dei greie for ordlaginga i språket, synkront og ofte diakront. Slike referanseverk er uomgjengelege som forskingsstøtte i dei fleste vitskapsgreiner.

3. Ordbøker og kunnskapsteori

Det har vore vanleg i nyare leksikografisk faglitteratur å hevda at å skriva ordbøker er eit handverk, eller ein kunst. Implikasjonen er at formatet for ordboka og sjølve ordartikkelen ikkje har eit fundament som tilfredsstillar vitskapsteoretiske krav, eller at det teoretiske fundamentet ikkje er viktig. Ei typisk utsegn (frå ei mykje brukt handbok) er denne:

³ Historisk-Filosofisk fakultet, Universitetet i Oslo avklara den vitskaplege statusen til leksikografien ved fakultetshandsaming i 1989 (Utvalget for forskingsspørsmål 1989).

Lexicography is a craft, a way for doing something useful. It is not a theoretical exercise to increase the sum of human knowledge (Landau 2001:153).

Eg er usamd med Landau; føremålet med den vitskapelege ordboka er sjølvstøtt å auka summen av menneskeleg kunnskap om dei systematiske samanhengane mellom einskildord og alle eigenskapar som ord har, frå lydleg realisering til tyding og språkbruk. Det finst vitskapsteori og modellar for korleis ordboksartiklar og ordbøker må vera om dei skal fylla føremålet.

For leksikografien er det uheldig at det i periodar har mangla ein eksplisitt fagdiskusjon om grunnleggjande teori og metode. Dette har gjort det lettare å avfeia leksikografi som eit handverk utan teoretisk fundament, og skapar vanskar for etableringa av leksikografi som akademisk disiplin.

3.1. Ordboka, ordboksartikkelen og tidleg vitskapsteori

Standardteorien for etablering av vitskapeleg sanning i mellomalderen og tidleg nytid bygde på Aristoteles si lære om logikk. Aristoteles drøfta aldri språk direkte, men han kjem inn på språkrelaterte emne i fleire avhandlingar om logikk og litteratur. Særleg viktig er den vesle teksten *Perihermeneias* (Arens 1984 10-11). Formatet for ordboksartikkelen kan reknast som ei utsegn om kva noko er, eller kva for vesentleg(e) eigenskap(ar) noko har, og slik sett innordnast under aristotelisk logikk for etablering av sanning. Ordboksartikkelen kan reknast som ei deklarativ utsegn om ekvivalens; i ein gjeven kontekst er oppslagsordet likeverdig med definisjonen (Bussmann et al 2006:283-4). Likningsforma er ein svært sterk reiskap for å etablere samheng mellom oppslagsord og omgrepsbeskriving, like sterk for språk som for naturvitskapane. Eit register av ordtilfanget i eit språk, i form av oppslagsord som er sette lik systematisk oppbygde definisjonar, kontrollerte mot gode kjelder, aukar summen av menneskeleg kunnskap om det aktuelle språket og om omgrepsverda som språket refererer til, og dermed kunnskapen om korleis verda ser ut gjennom dei brillene som det språket utstyrrer oss med.

Aristoteles er far til andre utsegner om språk som har vekt for leksikografien, endå om teksten handlar om filosofi eller litteratur. Han peikar på at ei ordform i isolasjon ikkje tyder noko (Arens 2000:369), at

grunnord skil seg frå samansette ord ved at dei ikkje kan analyserast med oppdeling (Bywater 1962:71), at bøygde former er variantar av substantiv og verb, at sambandet mellom ord og tyding er menneskeskapt og vane-bestemt (konvensjonelt) - altså ikkje gudegjeve (Arens 2000:367) og at forståing av ord, og dermed formidlinga av tyding, er avgrensa av individuell og sosial språkleg kompetanse (Bywater 1962:71). Alt Aristoteles seier om språk og logikk, føreset elles høg morsmålskompetanse.

Tekstane til Aristoteles, med kommentarar, var pensum langt inn på 1700-talet. Det har vore elementær vitskapsteori at sambandet mellom ord og tyding kan dragast ut av etablert (og dokumentert) språkleg praksis; at ordboksartikkelen skal setja oppslagsord (pluss bøygde former) lik ein eller fleire definisjonar (som kan skiljast frå kvarandre ved å sjå på språkbruk); og at spørsmålet om definering må sjåast i høve til mottakarbehov.

Premissane for forma på ordboksartikkelen var sjølvsgade så lenge aristotelisk logikk stod på pensum for folk med utdanning. Ein kan undrast på om dette er grunnen til at formatet på ordboksartikkelen – etter det eg veit - aldri er eksplisitt kommentert i tidlege ordbøker (sein mellomalder og renessanse). Ordboksartikkelen var ein kjend standard for korleis sanning om språk blir etablert. Formatet på ordartikkelen og ordboka må ha vore velkjend for alle studerte menneske i Norden, som hjelpemiddel i å læra seg latin og gresk, og som format for registrering av sambandet mellom teikn og tyding. Ordinnsamlinga som vart sett i gang av Matthias Moth, hadde ikkje vore mogleg om ikkje slik kunnskap var allemannseige⁴.

Den første ordboka med norsk ordtilfang er ei lita bygdemålsordliste med ord frå Sunnfjord, utarbeidd av soknepresten Christen Jenssøn og prenta i København i 1646 (Jenssøn 1646). Om ein berre ser på stad og emne, kan det vera uventa at denne ordboka er eit for si tid leksikografisk velordna produkt, med oppslagsord i grunnform, ordklasse, definisjonar og bruksdøme, og i somme tilfelle gloss på latin. Men ordboka er skriven av ein prest med universitetsutdanning. I studiet må han ha blitt vel kjend med ordboks-formatet og ordbokssjangrane i samtida. Alt Christen Jenssøn har gjort, er å bruka denne allmennkunnskapen på sitt eige heimemål, liksom Matthias Moth føresette at informantane hans skulle gjera, femti år seinare.

⁴ Sjå t.d. omtalen av Matthias Moths ordbok <dsl.dk/sprog/ordboger-og-sprogteknologi/moths-ordbog>, eller omtalen av dei norske bidraga i Grønvik 1998.

3.2. Ny tid og ny teori

Det finst nokre sjølvstendige ordbøker frå sein mellomalder, i manuskript, men den sjølvstendige ordboka som sjanger høyrer nytida til. Ordboksmanuskript er arbeidskrevjande å setja opp, men òg svært nyttige. Høvet til masseproduksjon av tekst, saman med lærehunger og nye krav til utdanning, skapte ein stor marknad for ordbøker på 1500-talet. Mange ordbøker vart trykte fleire gonger, med utbetringar og redaktørskifte. Samstundes som dei første store dokumentariske ordbøkene vart etablerte (*Vocabulario della Crusca* (VDC), *Dictionnaire d'Academie Française* (DAF)), voks det opp ein blømande marknad for kommersiell leksikografi.

Skjedde det teoriutvikling i høve til ordboksformatet i denne tida? Ja, ein ser ny teori, dels som følgje av generell vitskapsteoretisk utvikling, dels retta inn mot både makrostruktur og mikrostruktur i ordboksformatet.

Det generelle først: Dei store ordboksprosjekta frå 1500- og 1600-talet viser at leksikografien utvikla seg i takt med haldninga til vitskap generelt: Ordboksartikkelen kunne ikkje berre vera ei logisk korrekt utsegn om ekvivalens, han måtte omfatta prov, i form av kjeldefeste døme på verkeleg språkbruk. Det gav empirisk grunnlag for å hevda at oppslagsordet fanst, og korleis det var brukt. Denne utviklinga grunnjev arkiva som vart oppbygde for VDC og DAF.

Masseproduksjonen av trykt materiale førte også til ei anna endring som fekk følgjer for ordboksformatet. Som undervisningsemne trong skriftspråket fast form, særleg ortografisk, men og på andre område. Fast rettskriving var ein føresetnad for den mest populære makrostrukturen for ordbøker, alfabetisk artikkelordning.

Andre ordbokstekniske nyvinningar gjeld kategorisering og modellbygging, og knytter seg til det at kontrollert typografi kan brukast til å markera feltinndeling. Omtale av slike nyutviklingar finn ein i til dels føreord og brukarretteiingar, med vekt på nyvinningar i høve til konkurrentar⁵. Også denne måten å tenkja analytisk på, kan knytast til samtidas nytolking av aristotelisk metode (Steinfeld 1986:33).

Tida frå reformasjonen og fram til om lag 1800, står først og fremst i den normative leksikografiens teikn. VDC og ADF la seg i forkant ved å prøva å

⁵ Sjå t.d. Mooijaart 2013: 6-9, med ein gjennomgang av formatutvikling i eitt- og fleirspråklege ordbøker for hollandsk på femten- og sekstenhundretallet.

normera ein skriftspråksstandard i framvekst, og vart dermed sterkt preskriptive. I England kom det første store, litterære ordboksverket først i 1755, med Samuel Johnsons *Dictionary of the English Language*. Dette tiltaket kunne utnytta to hundre års litterær oppblomstring innanfor mange sjangrar. Normativiteten i Dr. Johnsons ordbok gjer seg derfor gjeldande i lemmaseleksjon, tilpassing av rettskrivinga og utval av døme frå eksisterande tekst, men ikkje i ope preskriptive reglar for korrekt uttrykksmåte⁶ (Brewer 2000:40-41).

3.3. Komparativ filologi og historisk leksikografi

Dei store vitskaplege morsmålsordbøkene som vart påbyrja på 1800-talet har sett ein standard for leksikografi internasjonalt. Dei best kjende i tillegg til dei nordiske er *Deutsches Wörterbuch* (DWB), *The New English Dictionary* (no: *The Oxford English Dictionary*, OED) *Svenska Akademiens Ordbok* (SAOB), *Ordbog over det Danske Sprog* (ODS), *Woordenboek der Nederlandsche Taal* (WNT). Desse ordboksverka tok sine mål frå den språkvitskaplege teoriutviklinga i samtida, og i opposisjon til den normative leksikografien 1700-talet. Utgangspunktet var framstega til den komparative språkvitskapen tidleg i hundreåret, som gjorde det mogleg å handsama språkutvikling på ein systematisk måte. Språkutviklinga blir dedusert frå form på og bruk av einskildord. Ordboksprosjekta var derfor veleigna til å syna fram språket som diakront og synkront system. Ordbøkene skulle vera deskriptive og upartiske, og i prinsippet opne for alt ordtilfang. Det vitskaplege låg i at beskrivinga skulle byggja på røynelege og kjeldefeste språkprov, systematisk ordna i tid. Ved å jamføra systematisk med eldre språksteg og andre språk kunne ein kartleggja ordlaging og tydingsutvikling nøyaktig, og avdekkja generelle mekanismar i det språket som var under lupa.

Tankegangen bak dei historiske ordboksprosjekta var den same som for mange andre vitskapsgreiner på 1800-talet: komparativ metode anvendt på materiale ordna etter ein tidsakse, skulle gjera det mogleg å beskriva ikkje berre korleis røyndomen såg ut, men kvifor han var som han var og måtte vera nettopp slik, etter dei lovene som gjeld for fagfeltet. Ei systematisk

⁶ Jf. Preface, der Dr. Johnson gjer greie for prinsipp for normering og utval av oppslagsord og døme, og definisjonsprinsipp.

utforsking av ordtilfanget i dokumentert bruk vart slik sett ein naudsynt føresetnad for å nå høgare abstraksjonssteg i alle andre, meir spesialiserte og abstrakte deldisiplinar i språkvitskapen, som (diakron og synkron) fonologi, morfologi, semantikk og syntaks. Parallellen til utviklinga av t. d. det vitenskapsteoretiske grunnlaget for faget geologi er slåande.

I teorien var dei historisk-komparative ordboksverka opne for ordtilfang frå talemål utanfor standardspråket, men orda måtte ut frå den tids teknologi skrivast ned og visast i bruk i skriftlege døme. Fonetikken var i si første byrjing, og det fanst ikkje standardar for talemålstranskripsjon. Ordtilfanget som skulle dokumenterast, måtte derfor vera dokumentert i skrift. Dermed vart dei første store vitskaplege ordboksverka tungt litterære. Når talemålsleksikografien kom, seinare på 1800-talet, var leksikografisk teori og metode mykje lik den som var nytta før. Utfordringane i talemålsleksikografien var dei same som for dialektologien, og knytt til metodar for isolering og organisering av lyd, transkripsjon og heimfesting. Eit tilleggsproblem var og er det at innsamling av talemålstilfang er kostar meir arbeid enn litterært tilfang, og er meir heterogent i form. Talemålsleksikografane har derfor ei særleg utfordring i identifikasjon av ukjende ordformer – er det nye oppslagsord, eller variantar av eit kjent ord? Slike spørsmål var dagleg kost for redaksjonen i NO, og har fordra omfattande morsmålskompetanse i redaksjonen frå både diakron og synkron synsvinkel.

4. Teknologi og ordboksutvikling

Ordboksproduksjonen har vore gjennom to store teknologiske revolusjonar: overgangen frå manuskript til trykte bøker og deretter frå manuskript som laupande tekst til manuskript i ein elektronisk mal med førehandsdefinerte grenser. Den første overgangen skapte langt på veg sjangeren ordbok som sjølvstendig produkt. Den andre overgangen har endra vilkåra for produksjon og tilgjenge, slik til at vitskaplege ordbøker kan bli folkeeige.

For ordboksprosjekt gjev dette tre innsatspunkt der overgang til IKT kan gje vesentlege vinstar. Det gjeld tilhøvet mellom råmateriale (kjeldetilfang) og produkt (ordboksverk, ordartikkel), produktformat (ordboka som heilskap og artikkelformat) og produksjonsprosedyre (delvis automatisert innskrivingsformat).

4.1. Følgjer av trykking for ordbokssjangeren

Den gjennomgåande forbetringa for kunnskapsformidling som trykkjekunsten førte med seg, gjeld visualisering av systemet både i makro- og mikrostruktur. Trykkjekunsten opna også for masseproduksjon av ordbøker, og kan seiast å ha skapt sjangeren. Utdannings- og kunnskapseksplorasjonen som prega perioden, kravde effektiv formidling gjennom fleire språk enn latin, og morsmålets stilling og prestisje vart ei omstridd sak gjennom heile 1600-talet, ikkje minst i Danmark (Steinfeld 1986:44 ff.).

Ordbokssjangeren var alt før 1600 utvikla med etter måten velordna format, i tett tilknytning til funksjon (Simpson 2007: 18). Her kunne morsmålsordboka byggja på meir enn 100 års tradisjon med trykte to- eller fleirspråklege ordbøker (Hanks 2010:12). Det tok ikkje lang tid før dei grunnleggjande prinsippa for korleis ei ordbok skulle sjå ut, var på plass: artikkelordning i kolonner, typografisk separering av kategoriar, alfabetisk eller tematisk ordning av ordartiklar.

Også mikrostrukturen er svært gjenkjenneleg sett frå moderne synsstad. Kategoriseringa i dei tidlege prenta ordbøkene er lett å sjå, takk vere typografiske verkemiddel som storleik, satstype og farge. Det at kategoriseringa blir så synleg, er til hjelp i sjå og skilja mellom ulike kategoriars eigenskapar, og i å auka konsistensen i gjennomføringa.

Ordboksartikkelen har truleg vore forstått som ein serie med skriftlege uttrykk for kategoriar så lenge sjangeren har eksistert, jf. ordbøker frå tidleg nytid (Elyot 1538; Maaler 1561; Cawdrey 1604). Ei rekkje kategoriar er ordna etter eit indre system der hovuddraga ligg fast frå ordbok til ordbok, samstundes som det er rikeleg med rom for tilleggsspesifisering. Det går nesten ikkje å skilja mellom eittspråkleg og tospråkleg - ein må heller sjå på kva for språk det er fokusert på, eventuelt gjort greie for, og kva for kategoriar som er viktigast i kvar ordbok.

Lat oss ta Cawdrey (1604) som døme: Han bruker utheva oppslagsord med liten forbokstav og i grunnform, og definisjonsformat med same setningsleddform som oppslaget. Han instruerer i alfabetbruk, og gjennomfører same redaksjonsmål i oppslagsord og definisjonstekst. Ordboka hans er ei ordbok for "hard words", så opphavsspråk er markerte med ein forkorting; "fr" for fransk, "gr" for gresk osv.

Artikkelformatet – altså mikrostrukturen - omkring 1600 framstår som temmeleg standardisert. I tillegg til oppslagsord i grunnform finn ein markering av genus med artikkel, bøyingsformer førte etter fast mønster, etymologi (opphavsspråk for importord), definisjon (med synonym eller antonym), bruksdøme eller sitat (øg med kjelde), og motsvar i anna språk (oftast latin). I somme tilfelle er bruksdøma ordna etter tyding, med kommentarar som syner tydingsnyanse. Redaksjonelle døme kan bli nytta for faste uttrykk, sitat for meir kontekstbunden språkbruk. Fraseologi kan vera skilt ut og kommentert. Den tekniske revolusjonen som trykkjekunsten var, gjorde ordboka til standardreiskap i utdanning for mange fleire enn før.

Overgangen frå handskrift til trykking inneber endringar som kan minna om overgangen frå trykking til IKT (Hanks 2010). Ordboksartikkelen har med andre ord vore ein opplagd kandidat for innpassing i eit databaseformat lenge før datamaskina fanst.

4.2. Grunnlagsmateriale, arkivteknikk og meir typografi

Ordboksprosjekta som var inspirert av den historisk-komparative språkvitskapen, har det sams at dei vart bore fram av entusiasme i startfasen, som sentrale nasjonale identitetsprosjekt (Mugglestone 2006:4), og at det ikkje vart utarbeidd realistiske prosjektplanar. Det vart laga detaljerte faglege planar for kva ordbøkene burde innehalda, og sett mål for total storleik av typen ”fire til seks store band”. Krav til underlagstilfang, manuskriptlengd (totalt og per artikkel), tids- og kostnadsstudiar osv., manglar. Fleire prosjekt kviler på dugnadstanken – friviljuge skal ta seg av ekserpering og setelutskrivning, men behovet for opplæring vart oversett i utgangspunktet (Murray 2001:171 ff.). Etter kvart voks forståinga for at ordinnsamling ikkje er trivielt arbeid; Arkiva etter dei norske tiltaka, Det Norske Litterære Ordboksverk og NO, syner tett brevsifte med friviljuge medarbeidarar. For NO vart det skriva eit detaljert, eige instruksjonshefte (Hellevik 1956).

Spørsmålet om oppbygging av materialarkiv, og arkivteknikk, vart også undervurdert. Samtlege av desse ordboksprosjekta tok innleiingsvis lett på arbeidsmengd, logistikk og kostnad ved å byggja opp ekserptsamlingar som var gode nok for vitskapleg leksikografisk arbeid, i det omfanget som

planane vitnar om. Det er diverre symptomatisk at ordbokshistoriske arbeid seier lite om arbeidet med materialsamling og tilrettelegging.

Men ekserptsamlingane skulle leggjast til rette for redigering, og takast vare på for innsyn og kontroll. Dermed tvinga arkivoppbygging seg fram. Ivar Aasen arbeidde aleine og skreiv lister; James Murray byrja med setlar, brukte hyller, og sette familien i arbeid; skuffer med arkivkort i fast format og vern mot nedstøving kom etter kvart, likeins sorteringskasser og andre tiltak til lette for redigeringa. I somme ordboksverk (t. d. OED og SAOB) vart det skipa egne avdelingar som førebudde tilfang til redaktørane.

Grunnlagsmaterialet skulle supplera informasjon om både form og innhald. Ambisjonsvoksteren for den heildekkjande nasjonale morsmåls-ordboka hadde gjort artikkelformatet mykje meir omfattande enn tidlegare, særleg formdelen som skulle gjera greie for utviklingshistoriske tilhøve. Tilfanget for kvart ord derfor måtte sorterast fleire gonger, og sorteringa – setelbunkar på bord – var sjølvsagt sårbar (Bø 1989:87 f.).

NO kom først i gang i 1929, og lærte av forgjengarane. Arkiva til NO vart laga i einsarta format og finalfabetisert i arkivskuffer frå første stund. Utfordringa i NO-tilfanget var såleis ikkje arkivteknisk. Men arkivutbygging vart etter det eg veit aldri rekna på som ei eksplisitt arbeidsoppgåve, og dermed til dømes aldri vege mot arbeidsinnsats i korpusoppbygging, før på 1990-talet. På 1990-talet vart papirarkivet fotografert og digitalisert, og det digitale Setelarkivet kom på nett i 1998. Ein heldt fram med produksjon av papirsetlar i ei tilleggsarkiv til 2001, men då var det slutt. I dei vitskaplege ordbøkene vart typografisk feltmarkering elles driven til nye høgder etter som ordartikkelen utvikla seg i kompleksitet. I dei nordiske vitskaplege ordbøkene er typografi som feltindikator utvikla lengst i SAOB; men alle store ordboksverk frå 1900-talet legg vekt på typografiske verkemiddel i formidlinga av leksikografisk innhald (sjå t.d. Osselton 2006:69).

Typografisk har NO aldri vore svært sofistisert; det var og er nytta éin font i ein storleik, i halvfeit, rettskrift og kursiv. I papirperioden var kategoriseringa av innhaldet dermed avhengig av rekkjefølgje og tekstmarkørar, meir enn av visuelle indikatorar.

4.3. Frå papirarkiv til digitale samlingar

Frå 1980 om lag braut datateknologien for alvor inn i ordboksverda, først som hjelp til meir effektiv setjing og prenting av manuskript, deretter som reiskap i strukturering av manus. Etter kvart kom det redigeringsformat som føresette førehandskategorisering av ordboksinnhaldet, med redigering rett inn i artikkelskjema, lagring av manus i (relasjons)databasar, og tekstvisning i fleire ulike format, alt etter kva medium det er tale om. Mitt første møte med dette arbeidet kom på eit arbeidsseminar i København i mars 1988, om innskrivingsapplikasjonen Compulexis, organisert av ODS-S og GDO. Skiljet mellom digital tilrettelegging av materiale og satsing på en rein innskrivingsapplikasjon vart lyft fram, og representerer framleis eit hovudskilje i tilnærminga til kva overgang til digital plattform inneber.

For NO kom digitalisering av tilfanget først, via Dokumentasjonsprosjektet i Noreg (jf Ore & Kristiansen 1998:19 ff.). NO dekkjer tale og skrift gjennom fire hundre år, og kjeldene speglar dette. Heterogeniteten i det skriftlege uttrykket er stor, og krev solid filologisk innsikt for at tilordninga frå ordform i kjelde til oppslagsord skal bli rett. Eit særleg problem for NO var rettskrivingsutviklinga frå Aasens tid til notid, med ein serie ortografiske revisjonar som m.a. innebar endringar i framlyd, og dermed spreining av tilfang til fleire bokstavar i alfabetet. Setelarkivet var éi kjelde, andre kjelder var normerte etter eldre rettskrivingar, eller unormerte. Det mangla både oversyn over omfanget og kunnskap om arbeidssparing ved betre tilrettelegging – og i så fall korleis ein best kunne gjera det.

NO-redaksjonen visste at normeringa i Setelarkivet var så mangearta at det seinka effektiv redigering. I denne situasjonen kom framlegget frå Christian-Emil Ore om å laga ein felles indeks for alle samlingskomponentar nærast som svar på ei bøn (Ore 2000 og 2001). Indeksen fekk namnet Metaordboka (MO), og dei første komponentane var Setelarkivet⁷, Trønderarkivet⁸ og Grunnmanuskriptet.⁹

⁷ Setelarkivet vart oppbygd med dugnadsinnsats frå 1929 og utover, no på 3.2 mill setlar, med både talemålsopplysningar og litteraturekserpt.

⁸ Setelarkiv oppbygd ved nordiskmiljøet ved Universitetet i Trondheim. Trønderarkivet dekkjer talemålet i Trøndelag og på Nordmøre, og har om lag 200 000 setlar.

⁹ Grunnmanuskriptet er ei første samarbeiding av sentrale kjelder for Norsk Ordbok, utført om lag 1930–1945 (Vikør & Wetås 2016: 19-20).

Overføringa av NO til digital plattform måtte byrja med rydding i underlagstilfanget, for utan rydding kunne vi ikkje proporsjonera dei attverande banda, og utan proporsjonering fekk vi ikkje lagt ein plan for framdrift. MO er ordna strengt alfabetisk, og er det beste utgangspunktet både for lemmaoversyn og for proporsjonering av heile ordboka. For proporsjonering av dei attståande banda var det viktig at tilfanget for kvart lemma var samla under ei einaste oppslagsform i MO. I det første prosjektåret gjekk 15 årsverk til normering av MO frå i til å. Artikkeltalet fall med om lag 15 prosent. Tilhøvet mellom grunnlagstilfanget og heile ordboksverket vart handtert ved at MO vart nytta til å proporsjonera dei attverande 8,5 banda (i-å). Alfabetfordeling på 12 band vart rekna ut for heile verket og for dei attverande banda, i bolkar på 20 sider. Dermed fanst det ein hovudplan som viste kor mykje plass ein hadde per alfabetbolk, og kor mykje som måtte gjerast per år om prosjektet skulle bli fullført innanfor tidsramma. Denne planen vart følgd strengt og heldt til og med band 11 og halvvegs inn i band 12.¹⁰

Tilrettelegging av ulike kjelder for meir effektiv bruk vart eit arbeidsområde i heile prosjektperioden. Den mest arbeidskrevjande oppgåva har utan tvil vore innordning av unormert tilfang (målføreordsamlingar og eldre ordsamlingar) i MO. Dette inneber at prosjektet også har måtta handtera retrodigitalisering av eldre tilfang, både utgjeve og uutgjeve.

Det var behov for utviding av grunnlagstilfanget. Nynorskkorpuset kom til i 2003 og har vokse under heile prosjektperioden¹¹. Det er allment kjent at manuelt ekserperte språksamlingar er gode på leksikalsk breidd, men svake på å spegla bruksfrekvens. Føremonen med tilfangsauken i Nynorskkorpuset viste seg i at ei mengd språklege fenomen kom i rettare lys, både dei frekvente og dei sjeldsynte, i samsvar med eller som avvik frå overordna mønster.

Sidan tilfanget voks gjennom heile prosjektperioden, vart finproporsjonering per band gjort rett før oppstart av nytt band, men forskuving mellom bolkar eller band vart ikkje gjord. Det var planen frå 2005 som galdt.

¹⁰ Som kjent vart slutfasen av prosjekt NO2014 forstyrra av konflikten med leiinga ved UiO. Dette er grundig drøfta i Vikør & Wetås 2016 og vil ikkje bli vidare omtala her.

¹¹ Nynorskkorpuset har no om lag 105 mill tokens.

Som del av den generelle proporsjoneringa vart artikkelomfang i tal linjer vekta etter materialmengd. MO-artiklar med eitt belegg – hapax - fekk 0 linjer. Dei største MO-artiklane fekk eit tak på 1260 linjer (10 spaltar).

Dette var ikkje bindande for redaktørane. Artiklar kunne gjerast større eller mindre enn den føreslegne ramma etter fagleg skjønn. Men den aktuelle alfabetbolken kunne ikkje gå utover den tildelte ramma, så sprengd ramme ein stad inneber kniping ein annan stad innanfor same alfabetbolk.

4.4. Ordboka frå manuskript til database

Overføring til digital plattform for eit pågåande prosjekt er ein balansegang. Ein vil ha føremonane med digitalisering, men ikkje skjera av sambandet med den predigitale fortida. For NO var hovudføremålet å få utnytta den sterkaste sida ved elektronisk datahandsaming, som er rask og lik handsaming av likearta data, og få redigeringa innordna i ein fast modell.

Nøkkelordet er ”likearta”. For å oppnå gevinst ved digitalisering, må dataa vera nøyaktig kategoriserte etter kriteri i faget og disiplinane som prosjektet spring ut av. Såleis vil digitalisering av eit gammalt prosjekt ofte innebera ei gjennomvasking av og opprydding i dei parameter som fagtradisjonen ber med seg – så også for NO, både i høve til den internasjonale disiplinen leksikografi og i høve til ulike sider av filologien og lingvistikken, anvendt på eit nordisk språk.

Ein premiss for digitalisering av gamle NO var at verket skulle framstå som ein einskap før og etter overføringa – dei fire første banda skulle hengja tydeleg saman med dei åtte siste. Dette innebar at redaksjonell rettskriving¹² og alfabetisering skulle stå uendra¹³, og at kategoriar og ordningsprinsipp skulle stå uendra så langt råd. Det innebar også sterk vekt på at papirordboka skulle tilfredsstilla ordboksestetiske krav. Dei fire første banda av NO var innskrivne som elektronisk tekst og korrekturlesne sist på 90-talet. Teksten vart analysert manuelt og datamaskinelt. Analysen synte at papirutgåva av NO var eit svært komplekst dokument.

NO skulle også få eit vitskapleg løft: IKT-overgangen skulle bidra til å framheva ordboksverket som ein gjennomført og riktig proporsjonert analyse

¹² 1938, tradisjonelle former

¹³ Endå om nynorskrettskrivinga er endra fleire gonger etter 1938, seinast i 2012.

av norsk ordtilfang. Kvar artikkel skulle vera ei lita vit-skapleg avhandling, med ei korrekt og systematisk gjennomført oppsummering av underlagstilfanget, i ei form som er klar og overtydande, og like brukande for lek og lærd. Uløyste flokar skulle isolerast og få ei løysing.

I ei vitenskapleg ordbok skal det vera tett samheng mellom underlagstilfang og ordartikkel, ned på tydingsnivå, og med rett rekkjefølgje på utvalde sitat. I ein leksikalsk ordbase kan dette gjerast med hyperlenking: brukarar av ordboksverket kan få høve til å sjå rett inn i underlagstilfanget. Slik kan brukaren sjå om ordboka svarer til sitt kunngjorde føremål, om ordboksartikkelen er dekkjande for tilfanget, og om tilfanget treng supplering for å få med udokumenterte nyutviklingar. Slik hyperlenking gjer også kvalitetskontrollen enklare innanfor redaksjonen – kollegalesaren kan klikka på ei lenke i staden for å gå og leita fram setlar i papirarkiv.

Innskriving av ordartikkelen gjekk med digitaliseringa over frå å vera ein komposisjon i fri tekst, til å bli utfylling av eit skjema, med tydeleg skilje mellom fritekst og bunden tekst (Grønvik 2005b:5). I ei ideell verd skal det å fylla ut eit skjema på nettet vera så lett at framgangsmåten verkar sjølvforklarande. Gjennom kartlegginga av arbeidsmåten til dei gamle NO-redaktørane var det kjent korleis dei arbeidde, og kva dei la vekt på. Strukturen i ordartikkelen var uendra i hovuddrag, men stramma inn og formalisert på mange punkt. Redigeringsformatet vart lagt opp tett opp til eksisterande arbeidspraksis, med dei tillempingane som følgjer av omleggingane som er nemnde ovanfor. Kjedefesting, teiknsetjing osv. er knytt til format og menyar. Kjelder og anna støttetilfang er så langt råd plassert i eigne databasar eller lister som så er kopla til redigeringsapplikasjonen. Det er høve til å kopla artiklar til kvarandre på artikkel- og tydingsnivå. Detaljane i korleis ein arbeider innanfor dei ulike applikasjonane er det gjort greie for i brukarmanualar, medan den leksikografiske grunngevinga, knytt til konvensjonane for NO, er å finna i Redigeringshandboka (Grønvik & Gundersen 2015).¹⁴

NO blir lagra i ein relasjonsdatabase der innhaldet kan uttrykkjast som pdf etter fast stilark (prenteklar sats), ordartikkelformat på nett, eller i ulike

¹⁴ Redigeringshandboka fekk i hovudtrekk form 2003-2007. Seinare tillegg er innarbeidde laupande, fram til nettpublisering 2015.

format innanfor grensesnittet til den leksikalske databasen¹⁵. Alle eldre versjonar av artiklar blir logga og lagra.

Redigeringsapplikasjonen med tilliggjande verktøy vart etter kvart utvikla til eit produksjonsanlegg for ordartiklar som dekkjer heile prosessen frå artikkeloppretting til publiseringsklart produkt. Den gjer det mogleg å a) generera artiklar frå indeksert tilfang, med lenke til tilfanget, b) annotera og sortera tilfanget og lagra analysen, c) generera artikkelhovudet frå artikkelhovudet i Metaordboka, d) lenka tilfanget for kvar definisjon direkte til definisjonen, e) følgja produksjonsflyten per artikkel og for ordboksmanuskriptet som heilskap. Det er også lagt til rette for langtidslagring av ordboksmanuskriptet.

4.5. Frå einskildmedium til multimedium

Det er ikkje urimeleg å rekna med at store ordbøker i framtida kan bli del av informasjonsnettverk der lyd, bilete og andre informasjonkjelder heng saman med ordboka (Grønvik & Ore 2016:139). Her skal eg halda meg til det som er gjort og har vore påtenkt for NO.

Indre hyperlenking. NO på digital plattform fekk høve til hyperlenking mellom artiklar og definisjonar. Dette kan ein vel rekna som standardutstyr etter kvart. Føremålet var mellom anna å hindra t.d. eittordsdefinisjonar som ikkje fanst som oppslagsord, men og generelt å gje høve til den presisjonen definisjonsspråket som ei vitskapleg ordbok bør ha.

Ein annan type indre hyperlenking går til samansette ord som ikkje er redigerte, men blir tekne med som døme på ordlagingspotensialet til oppslagsordet. Her blir det oppretta ein artikkel i databasen, som i utgangspunktet er innhaldstom. Men tilfanget finst, og om ein seinare kjem til at ordet fortener å definerast, så er det mogleg.

Ytre hyperlenking. I og med at alle NO-artiklar er genererte frå Metaordboka, er det umogleg å oppretta artiklar utan underlagstilfang. Ein kan spesifisera dette vidare ved å oppretta direkte kopling frå definisjon eller sitat til det relevante belegget. Vidare er det mogleg å oppretta sorterar til kvar artikkel. Då hentar ein inn tilfanget frå Metaordboka og eventuelt frå Nynorskorpuset, og annoterer, sorterar og ordnar etter behov. Sorteraren er

¹⁵ artikkeltre, innskrivingsskjema, laupande tekst.

ei viktig stø for redaktørane i arbeidet med store artiklar, men også små artiklar kan ha bruk for annotering av tilfanget. I sorteraren blir annoteringa lagra, slik at ho er tilgjengeleg for framtidig revisjon.

Det har vore planar om å lenka NO (via MO) til målførekorpuset ScandiaSyn, til dømes for å få lyd til sitat. Til dette har vi mangla finansiering, men det som tel meir, er at det ligg personvernrestriksjonar på bruk av tale. Fleire mindre nettbaserte målføreordbøker har innlest tale til oppslagsord og sitat (t. d. Skolt & Wøllo 2013), så dette er noko som kjem.

Lenking til ikkje-leksikalske resursar. Den siste forma for hyperlenking som det vart tid til for NO, var lenking til kart, for å syna målføretilfanget per ordform, definisjon og sitat. Dette vart mogleg av to grunnar – geografikjeldene til NO låg føre i ein hierarkist og vekta database, og Kartverket fristilte sine kartdata frå 1.1.2014, i eit kart med kommunegrensene frå 1940-talet, då heimfestingssystemet for NO vart fastsett. Karta i nettversjonen av NO er populære, og har utløyyst respons frå brukarar i form av tilleggsopplysningar til ordartiklar.

5. Kva måtte endrast i ei elektronisk NO?

Til no har det handla om tilhøvet mellom NO og underlagstilfanget. Men kva måtte endrast i sjølve NO? Det var det fire område som særleg kravde tiltak. To galdt primært makrostrukturen, to mikrostrukturen. For makrostrukturen var det tilordning av funksjon til ulike artikkeltypar, og sams, streng alfabetisering og avsnittsordning (sjå del 6 Makrostruktur). For mikrostrukturen vart det viktigast å setja ein standard for det NO kallar normalartikkelen, og ei standardløyising for handtering av fleirordige uttrykk (sjå del 7 Mikrostruktur).

Ei overordna endring var innføring av ein standard på tilfanget for eit redigeringsverdig¹⁶ oppslagsord (Grønvik 2005a:157). Før omlegginga til digital plattform hadde parolen vore at alt tilfang skulle utnyttast. Det førte til at NO 1-4 omfatta ufullstendige artiklar bygd på få belegg og utan dei opplysningskategoriane ein fullstendig ordartikkel krev. Denne framgangsmåten vart erstatta med grunnkrav til materialet attom kvar artikkel. Krava

¹⁶ Kommenter tilhøvet mellom redigeringsverdig og leksikalisering under vitenskapelig vinst

var a) oppslagsordet måtte vera sikkert identifisert med ordklasse, i systemet for ordlaging og ordhistorie. b) ordet måtte vera godt nok dekt i kjeldene til at det kunne skrivast ein definisjon med minst ei godkjend kjelde (Grønvik & Gundersen 2014:40).

Krava vart realiserte i redigeringsapplikasjonen slik: Normalartiklar må genererast frå ein MO-artikkel med belegg frå redigeringsverdige kjelder. Eit framlegg om artikkelstorleik, basert på materialmengd, følger med den oppgenerte ordartikkelen. Tilfanget følger med til sorteraren, der redaktøren sorterer tilfanget etter eigne kriterium.

Når ei ordbok skal over på digital plattform, må ein ha ein modell for ordartikkelen. Det skal vera ein generell modell som dekkjer alle ordartiklar. Då er det to moglege utgangspunkt; ein kan tenkja seg ein modell for ein pårekna gjennomsnittsartikkel, ut frå ei førestelling om korleis den bør sjå ut. Så må ein tillata påbygging for dei artiklane som krev annleis handtering. Det andre utgangspunktet er å laga ein maksimumsmodell som dekkjer dei største og mest komplekse ordartiklane, og så bruka berre delar av dette formatet på dei fleste ordartiklar.

For NO vart båe metodar prøvt. Det var alt laga eit digitalt format for *Bokmålsordboka* og *Nynorskordboka*. Vi eksperimenterte med utviding av det formatet det første prosjektåret. Men behova for artikkel-formatet i NO sprengde modellen. Hausten 2002 vart artikkelformatet nyskrive som maksimumsformat, med så eksplisitt kategorisering som det var rasjonelt å gjennomføra, gjeve at ordboka skulle framstå som ein samla heilskap før og etter innføringa av redigering på digital plattform.

6. Endringar i makrostrukturen

Utval og ordning av ordartiklar i ei ordbok, – makrostrukturen – byggjer på behovet for å lausriva gloseopplysningar frå tekst og organisera innhaldet i ei generell form. Den første og største endringa kom med trykkjekunsten, i overgangen frå glossar som følgjer ein tekst, til separate dokument med eigne ordningsprinsipp. Med sjølvstendig makrostruktur blir ordboka ein fleirbruksgjenstand. Dei to hovudordningsprinsippa for sjølvstendige ordbøker - alfabetisk eller tematisk ordning av oppslagsord - fanst før den trykte ordboka, men vart standardiserte som følgje av masseproduksjonen.

Andre makrostrukturelle løysingar er tematisk gruppering av artiklar, eller gruppering etter ordklasse. Men uansett artikkelordning er den trykte ordboka eit samhengande dokument. Ho kan lesast som laupande tekst, der lesaren alltid ser to heile boksider.

Leksikalske databasar skil seg frå papirordbøker ved at den einskilde ordartikkelen må kunna stå aleine. Han inngår i ei artikkelliste som blir vist fram, men artikkellista er bestemt av søket i databasen og varierer dermed frå gong til gong. Ein redaktør kan derfor ikkje bruka definisjonsplassen til å visa til føregåande artikkel, slik papirordbøker kan ha døme på. NO var redigert innanfor tradisjonell god skikk for vitskaplege ordboksverk, men var tenkt som laupande tekst. Derfor kunne ein finna på definisjonsplass ”jfr. flg”; som tyder noko slikt som ‘sjå på tilhøvet mellom neste oppslagsord og tilhøyrande definisjon, så forstår du kva dette ordet tyder’. I ein leksikalsk database, der kvar ordartikkel skal stå for seg, er dette ikkje godtakande.

6.1. Alfabetisk ordning med avvik

Den historisk-komparative språkvitskapen har som ein av sine følgjedisiplinar læra om ordlaging, særleg om avleiing. I etymologiske ordbøker kan tilfanget ordnast med oppslag på rot eller det ordet som er nærast rota, og med avleiingar og samansetningar knytt til same ordgruppe. Ei slik ordning bryt ofte med streng alfabetisk ordning for alt ordtilfang.

Noko av den same måten å ordna ordtilfang på, har flote over i dei større vitskaplege ordbøkene. Det gjeld særleg handteringa av samansetningar. Samansetningar med tilleggsopplysningar som står innbakte i andre artiklar, får uklar status: er dei samansetningsdøme eller er dei egne artiklar?

Ei elektronisk ordbok kan ikkje ha både alfabetisk og ikkje-alfabetisk sortering av artiklar. I NO vart dilemmaet løyst for det første ved strengt alfabetisk artikkelsortering etter første oppslagsform (somme artiklar har meir enn ei). Vidare skil NO mellom fire artikkeltypar: normalartikkel, tilvisingsartikkel, førsteleddsartikkel og etterleddsartikkel. Av desse kan berre normalartikkelen ha bruksdøme og sitat. Tilvisingsartikkelen peikar til ein eller fleire andre artiklar og kan ikkje innehalda anna enn slik tilvising.

Førsteleddsartikkelen og etterleddsartikkelen er skapte for å gjera greie for ordlagingssystemet i norsk på ein ordna måte. Etterleddsartikkelen kan berre brukast på ledd som ikkje finst som sjølvstendige ordformer (t. d. *-leis*,

adv, eller *-æring* m). Førsteleddsartikkelen har to undergrupper, dei tradisjonelle prefiksa (typen *mis-*, *van-*), og oppslagsord (med eller utan tilføyd fuge) som førekjem som førsteledd i samansetningar (typen *koke-*, *vass-*). Den siste undergruppa har to føremål; det første er å få gjeve formopplysningar som særleg vedkjem førsteleddet, for seg; det andre er å gjeva ei ordna framstilling av samansetningar til det aktuelle førsteleddet, vekk frå hovudartikkelen. Då blir samansetningane lista opp som døme, og som peikarar til tilfang i MO, utan tilleggsopplysningar.

7. Endringar i mikrostrukturen

Kva for endring har bruken av IKT ført til når det gjeld mikroformatet i ordbøker? Skal ein svara på det, må ein sjå på kva det betyr for leksikografien som fag at det blir laga modellar som simulerer eit generelt mikroformat for alle artiklar i ei ordbok, uavhengig av eigenskapane til oppslagsordet, og utan omsyn til tilfanget ordartikkelen skal byggja på.

Poenget med ein ordboksartikkel er å etablera ekvivalens mellom oppslag og definisjon. For å gjera denne likninga eintydig, kom det tidleg til spesifiseringar på båe sider av "likskapsteiknet", til dømes blir opplysning om ordklasse eller etymologi brukt til å grunnge skilje mellom homografar. På definisjonssida skil ein tydingar med definisjonar underbygde av bruksdøme. Idealet har vore stringens og konsistens, anten definisjonen har hatt setningsform eller har vore sett opp som hyperonym med spesifiserande tillegg, supplert med synonym.

7.1. Artikkelskjelett og kategoriordning

Formatet for normalartikkelen i NO er planlagt som eit maksimumsformat for dei største og mest innfløkte orda. Det skulle vera like brukbart for store og små artiklar, og tilby same funksjonalitet for ordartiklar som kravde stor plass og dei som ikkje gjorde det, men som kunne ha like finurlege kategoriseringsbehov.

Gamle NO hadde ein til dels svært innfløkt artikkelstruktur. Det var ikkje alltid klart kva nivå ein var på i høve til den innleiande definisjonen. Fleirordsuttrykk og avleiingar kunne stå som innbakte artiklar i artikkelen,

og fleirordsuttrykk kunne bli behandla på fleire punkt i hovudartikkelen, eller i fleire artiklar, noko som er uønskt.

Det vart derfor gjennomført ei forenkling av det moglege artikkelskjelettet i normalartikkelen, for å letta redigering, kvalitetskontroll og ordboksbruk. Ordartikkelen fekk færre nivå, fastare ordning av element, og kjeldefesting knytt til menyar. Om ein tenkjer seg artikkelskjelettet som eit tre, ville ein berre kunne gjera full bruk av alle artikkelement i den eininga som ligg ytst på kvar grein, som i NOs (interne) metaspråk fekk namnet ”tydingsbolken”. Overnivå til tydingsbolkar kan brukast til gruppering og felles kommentar til fleire tydingsbolkar. Innanfor tydingsbolken vart det innført fast rekkjefølgje av element: definisjon, døme, undertyding, fleirordige oppslag (”underartikkel”), samansetningsrekkje. Kvar tydingsbolk kan ha éin hovuddefinisjon. Neste hovuddefinisjon måtte ha nytt tydingsnummer. Denne måten å redigera på gav fleire nummererte tydingar, grunnare struktur per tyding, og kortare tydingsbolkar. Det vart også sett reglar for mengd og utval av sitat, krav om at prototypiske døme måtte koma først osv. Dette høyrer ikkje så radikalt ut no, men i 2002 var det eit merkbart, men naudsynt brot med 55 års praksis.

7.2. Formopplysningane i stram orden

I NO blir oppslagsdelen følgt av den såkalla hakeparentesen, altså den delen av ordboksartikkelen som inneheld informasjon om uttale, kjelder før 1900, skriftleg og munnleg formvariasjon, og etymologi med eit bilete, ordets id-kort. Denne artikkeldelen krev ofte store målførekunnskapar og solid innsikt i den indre språkhistoria, pluss brei kjeldekunnskap. I NO var arbeidet organisert slik at kvar redaktør har redigeringsansvar for heile ordartikkelen – vi har ikkje eigne etymologispesialistar. Redigeringa av hakeparentesen måtte derfor leggjast til rette så ho vart mest mogleg systematisk og homogen, og gjennomførbar for nye redaktørar med mindre røynsle i slikt arbeid.

Maksimumsmodellen for artikkelskjemaet tok utgangspunkt i ein utopi, nemleg full dekning av former per oppslagsordord for alle kategoriar; døme frå alle eldre kjelder, full informasjon om alle rettskrivingsendingar, målføredøme i alle bøyingsformer og frå alle kommunar, omfattande etymologiske opplysningar om band til andre ord i eige språk og samband til

andre ordartiklar i NO, med litterære referansar osv. Så langt råd var innhaldet bunde til tabellformat med fast rekkjefølgje for element. Kjelde-
menyen var avgrensa og heimfestingssystemet gjennomarbeidd og vekta.

Dersom NO hadde vore redigert berre med tanke på digital utgåve, ville tabelltanken ha blitt trekt lenger; som det var, måtte vi t.d. godta forkorta målføreformer, som gjer ordformene svært vanskelege å søkja i.¹⁷ Det var bruk for tydelege reglar om utval og føring, og for eit fritekstfelt til etymologiske opplysningar som ikkje kan stykkjast opp i tabellform.

7.3. Eige artikkelskjema for faste fleirordsuttrykk

Fleirordsuttrykk har alltid vore eit vanskeleg felt i norsk leksikografi. Ein grunn var mangelen på oversikt. Ein annan grunn var tradisjonen for å knyta tydinga til hovudordet i fleirordsuttrykk. Det kunne føra til at same ordsamband vart handtert under fleire definisjonar i same ordartikkel. Vi la derfor opp til at leksikaliserte, fleirordige uttrykk skulle redigerast på lik linje med andre oppslagsord, i eit noko redusert artikkelformat som fekk namnet "underartikkelen". Underartikkelen står i ordartikkelen for hovudordet i fleirordsuttrykket. I store ordartiklar med mange og sjølvstendige fleirordsuttrykk kan ein leggja underartiklane under eigne tydingsnummer.

Sjølv om papirordboka var den viktigaste bestillinga då NO2014 kom i gang, var det påtenkt at det måtte koma ei søkbar utgåve på nett. Med fleirordsuttrykka i eigne underartiklar er det mogleg å søkja etter faste fleirordsuttrykk på same måte som etter vanlege oppslagsord.

Grunnkrava til ordboksføring av eit fleirordsuttrykk var at det skulle ha fast form, minst ei tyding som var uføreseieleg ut frå einskildorda og form som eit setningsledd, ikkje som ei setning. Det vist seg at frekvens spelar ei rolle når ein skal velja handtering. For redigeringa av fleirordsuttrykka har det vore uheldig med strengt alfabetisk redigering. Det beste innordningspunktet for eit fleirordsuttrykk kan vera ein ordartikkel i alfabetstrekket a - h, endå om tilfanget for uttrykket låg under eit ord lenger ut i alfabetet.

Diskusjonen om korleis ein best avgrensar leksikaliserte ordsamband gått att og fram mellom redaktørane i heile prosjektperioden. Krava vart ikkje

¹⁷ Sjå t. d. artikkelen "trettan talord", der målføreformene *trettan*, *trettån*, *tretta*, *trettå*, *tretto* er attgjevne som *-an*, *-ån*, *-a*, *-å*, *-o*. Denne føringa gjev at søk på t. d. målføreforma *trettå* ikkje vil gje tilslag.

definert veldig strengt innleiingsvis. Det var det ikkje grunnlag for, korkje i leksikografisk tradisjon eller i faglitteraturen på det tidspunktet. I staden leit ein på at det redaksjonelle skjønnet kombinert med kollegalesing og kvalitetskontroll ville føra fram til felles normer og felles praksis, som så kan studerast og gjerast greie for. NO inneheld no om lag 15000 faste fleirordsuttrykk i underartikkel, og spørsmålet om faste ordsamband i norsk er dermed i det minste tilrettelagd for vidare gransking.

8. Kvalitetskontroll

Det å ha ordboksmanuskriptet i ein strengt kategorisert leksikalsk database har vist seg å ha store føremonar både i førebygging og retting av feil. Når mange menneske skal arbeida med eit mangslunge tilfang, må det oppstå uvisse på ymse punkt, uansett kor god rettleiing ein har tilgjenge til. At ein då kan søkja opp den aktuelle kategorien og sjå på praksis hos andre, er ei støtte både i redigering og kvalitetskontroll. Slik feilsøking blir ofte utgangspunkt for redaksjonelle diskusjonar og utfylling av redaksjonelle retningslinjer. I somme tilfelle vart målretta utfylling av kjeldetilfanget ei følgje, såleis vart det lagt vekt på å få inn målføretilfang frå Nord-Noreg, som var underdekt innleiingsvis. Ei anna følgje var arbeid for å klårleggja tvilsame punkt når det galdt somme av 1600-talskjeldene og 1700-talskjeldene.

Eit hovudspørsmål i store prosjekt er dei interne arbeidstilhøva. Dei skal ikkje omtalast her, utover den generelle kommentaren at eit stort prosjekt må ha ei fagleg og administrativ leiing som kan og vil ta avgjerder når det trengst. Alle som har følgd NO-prosjektet veit at prosjektet har vore velsigna med svært dugande prosjektdirektørar og ein lojal og dyktig stab som har tolt ein krevjande arbeidssituasjon i lang tid. Eg har før jamført ferdigstillinga av NO med å gå Besseggen 15 år i trekk, og det er ein metafor eg står ved. Korkje tidsramma eller økonomien tillét feilskjer.

Systematisk utdanning av nye redaktørar kom i gang frå 2005. Den teoretiske opplæringa følgde Redigeringshandboka og redigeringsapplikasjonen ganske tett, som ein systematisk modulbasert gjennomgang av tilfang, prinsipp og handsaming. Denne opplæringa bygde på røyntslene frå dei første gruppene med tilsette. Undervisningstilfanget vart oppdatert i detaljar, men heldt seg stabilt i dei åra opplæringa gjekk føre seg, til og med 2014.

Overgangen til redigering på digital plattform var størst for dei redaktørane som var med frå gamle redaksjonen. Det var mykje nytt å setja seg inn i på detaljplan, men den største endringa galdt rett og slett det å godta andres innsyn i eige arbeid. Det vart redigert inn i ein leksikalsk database der alle kunne sjå ein i korta, ned til materialsortering. Det vart umogleg å halda oppe idiosynkratiske arbeidsmåtar. Dette var noko alle forstod og godtok, men ein slik arbeidssituasjon er ikkje den vanlege innanfor humaniora, og det seier mykje om lagkjensla i NO-prosjektet at eit fleirtal av dei eldre redaktørane innretta seg greitt etter den nye tid.

9. Leksikografi og samfunnsrelevans

9.1. Vitskapleg leksikografi som prov på eksistens

Dei første vitskaplege morsmålsordbøkene må reknast som moderniserings-tiltak for heile samfunn. Dei sette standarden for skriftspråk og la til rette for meir effektiv kommunikasjon og kunnskapsformidling, i erkjenninga av at latin ikkje er for alle. Utdanningsvesen og skriftleg offentlegheit var under utbygging, og dei vitskaplege ordboksverka ytte sitt når det gjeld etablering av ein standard. Dette gjer dei etter si tids beste praksis, i respekt for aristotelisk logikk kombinert med dokumentasjon frå røyndleg språk. Ønsket om effektivitet i kommunikasjon og bruk pregar også den tidlege formateringa av ordboksartikkelen.

Tida etter reformasjonen er ei tid med nasjonsdanning, statsbygging og kamp om revir. Det var ikkje tilfeldig at kardinal Richelieu støtta opp om Academie Francaise. Det ligg meir enn to hundreår mellom starttidspunktet for VDC og DWB, men båe vart til i ei tid då desse to ordbøkene i sine respektive språksamfunn kunne stå som symbol på eit folkefellesskap som oversteig styringsform – korkje Italia eller Tyskland fanst som statar då ordboksprosjekta vart lanserte. Den patriotiske motivasjonen attom dei andre vitskaplege ordboksverka som vart påbyrja på i tida fram til andre

verdskrigen er heller inne vanskeleg å få auga på, sjølv om uttrykket er smålåte i dei nordiske tiltaka.¹⁸

Det den store vitskaplege morsmålsordboka gjer for eit språksamfunn, er å dokumentere eksistens over tid, pluss mangfald, omgrepssystem og sjølvbilete for det språket, den kulturen og det samfunnet som ordboka spring ut av, på den måten som kvart verk spesifiserer. Ordboksverka utgjer ein svært synleg kulturell og intellektuell kapital. Slik sett får dei vitskaplege ordboksverka symbolfunksjon for heile språksamfunn, i tillegg til praktisk verd for brukarane.

Ei morsmålsordbok for eit gjeve språk kan sjåast som ei beskriving av eit lukka system. Sjølv om ein unngår alle teikn til sirkeldefinisjonar i ordboksteksten, utgjer ordtilfanget i ordboka teikninventaret for kultursfæren som dette eine standardspråket rommar. Det som motverkar lukking, er særleg to tilhøve: den uavgrensa evna menneska har til omgreps- og orddanning, og grenseflata mellom språket i ordboka, og andre språk og språksteg. Også her er likningsforma reiskap; i fleirspråklege ordbøker er forventninga om ekvivalens mellom oppslagsord og motsvar på målspråket like grunnleggjande som forventninga om ekvivalens mellom oppslagsord og definisjon i den eittspråklege ordboka.

Autoriteten til ei ordbok, og dermed til redaksjonen, kviler på at brukarane er samde i at ordartikkel-likningane er rette, og at samanhengen mellom dei er truverdig framstilt. Provsgrunlaget er bruksdøma som utdjuapar definisjonen. Morsmålsordbøker har òg ein føresetnad om konsensus mellom ordboksredaksjon og brukarar; redaksjonens oppgåve er å skilja ut det vesentlege ved omgrepsinnhald i og bruk av alle oppslagsord, og uttykkja dette på ein måte som brukarane kjenner att og går god for.

Dei vitskaplege morsmålsordbøkene er kraftfulle uttrykk for at ikkje berre eit språk, men ein språksamfunn, eksisterer, har skapt ein omfattande skriftkultur, og har intellektuelt overskot til å reflektera over seg sjølv og sitt eige ved å definera heile si omgrepsverd på eigne premissar. Når dette er fullført, kan ein ikkje sjå bort frå resultatet. Det blir ein grunnvoll for vidareutvikling og sikkert omdefinering i mangt, men ingen kan lenger

¹⁸ ”Allermest haaber jeg dog, at ordbogen maa blive et værk, der i vort land maa fremme kendskabet til vort sprog, og uden for landets grænser maa kunne tjene som hjælpemiddel for dem, der søger kundskab om vort modersmaal og vort aandsliv.” Verner Dahlerup, føreordet til første band av ODS (s. IX).

trekkja i tvil korleis dette språket ser ut, eller kva slags sjølvportrett dette skriftsamfunnet har levert.

9.2. Symbolverdi

Symbolverdien til dei vitskaplege ordboksverka vart og blir sett og forstått i andre språksamfunn som manglar vern frå ein nasjonalstat. Eit allment godteke ordboksverk kan lyfta opp underkjende talemål og private skriftmålsuttrykk til utkast til standardspråk. Ivar Aasen hadde eit standardisert bruksspråk som eksplisitt føremål for 1873-ordboka. Både NRO og NO høyrer til i denne forståingsramma, og eittspråklege ordbøker som skriftfester ikkje-standardiserte språk fører denne ambisjonen vidare.

Brillene som desse ordboksverka blir sedde gjennom, skiftar etter om det er eins eige morsmål som er dokumentert, eller eit anna. For morsmålsbrukaren er ordboksverket i utgangspunktet tenkt som eit skattkammer der ein kan finna det kollektive minnet til sitt eige samfunn og omgrepsapparat, uttrykt i sitt eige mål. For utlendingar fungerer dei store ordboksverka over morsmål som sivilisasjonsmarkørar for språksamfunnet som ordboka spring ut av. Såleis har det vore ein vanleg påstand at engelsk er det rikaste språket i verda. Del av grunnlaget for påstanden er ordboka OED. Det står også glans av DWB, SAOB, ODS og WNT. Tilsvarande finst dei ei utbreidd førestelling (som leksikografar og lingvistar gjer sitt beste for å avliva) om at språksamfunn utan store ordboksverk er fattige – forstått som ‘med lite ordtilfang og lita omgrepsverd’ (Fjeld 2008).

9.3. Legitimitet

Den viktigaste kjelde til legitimitet for dei vitskaplege ordboksverka, er at folk går god for at det biletet av språket som ordboksverket gjev, er sannferdig og rett. Det tyder i praksis at brukarane er samde i det dei sjølve kan kontrollera som morsmålsbrukarar, og godtek det ordboka seier som brukarane ikkje visste frå før. Denne folkelege tilliten er svært viktig for NO, ikkje minst av di så mange har vore med på å samla tilfang til språksamlingane gjennom så lang tid.

Noko anna er politisk legitimitet. Kristin Bakken gjorde i 2005 ein analyse av kva for legitimeringsargument ein kunne nytta for å få

oppslutning om fleire nordiske ordboksverk, irekna dei norske (Bakken 2006:17). For NO sin del nemner ho det særskilte offentlege ansvaret for nynorsk skriftkultur, forankringa i talemålet over heile landet, og tradisjonsaspektet. Eg trur eit viktig argument i tillegg no er at NO faktisk vart ferdig, har fått god mottaking, og blir brukt. NO og språksamlingane har også greidd å legitimera seg vitskapleg, i alle fall førebels, i og med at Universitetet i Bergen har teke på seg vidareføring i samarbeid med høgskulen i Volda, og den første stipendiatstillinga er lyst ut.

9.4. Tilgjenge til dei vitskaplege ordboksverka

Så lenge dei vitskaplege ordboksverka berre fanst på papir, var dei òg symbol på lærd eksklusivitet. Men den tida er forbi. No som dei vitskaplege ordboksverka finst i gratis nettutgåver kan alle bruka dei når som helst, via datamaskin eller mobiltelefon. Nettutgåva av NO vart lansert 15.3.2012, og nettsida vart straks ein viktig formidlingskanal. Eg kjenner ikkje brukarstatistikken for dei andre nordiske ordboksverka – alle er på nett og fleire har tilpassa applikasjonar for mobilbruk – og har ikkje tal for NO. Men publikumskontakten inn mot ordboksredaksjonen viser at NO når langt fleire via nettet enn via papirbanda. Særskilt registrerer vi interesse for talemålsdokumentasjonen og dei dynamiske karta, med utbreiing for ordformer og tydingar. Karta synleggjer både funn og lakunar i materialet. Dei tek vekk illusjonen om at språksamlingane har full dekning, og kan inspirera informantar til å bidra med ny informasjon. Gjeve den sterke interessa for lokalt talemål som finst i Noreg, vil ei god publikumsteneste både styrka NO vitskapleg, og auka legitimeringskrafta.

9.5. Den praktisk funksjonen til vitskaplege ordboksverk

Alle dei nordiske landa har ei etter måten lang ordbokssoge. Vi har velutvikla utdanningssystem og høg aksept for at skriftspråket som blir nytta i det offentlege rommet skal vera normert.

Om vi tenkjer kontrafaktisk – korleis står det til å språksamfunn der morsmåla ikkje er uttrykt i ei standardrettskriving, og der semja om samanhengen mellom ordform og tyding ikkje er dokumentert i ei stor morsmålsordbok? Hovudregelen er at morsmåla ikkje er del av den

eksplisitte infrastrukturen i slike samfunn. Mange er tidlegare koloniland, der kolonispråka framleis held stillinga som administrasjons- og skulemål. Mangelen på morsmål i standardisert form held samfunnsutviklinga attende. Krefter og kreativitet som skulle vore nytta til nyvinning, blir i staden brukte til å læra utanboks kva kjende og ukjende omgrep heiter på eit framandt mål som til dels ligg svært langt unna morsmålet.

Føresetnaden for at eit språk kan vera eksamensemne, er at det finst ein fasit for korleis morsmålet skal sjå ut i skrift, i form av vitskapleg språkdokumentasjon. Dei viktigaste komponentane er ein grammatikk og ei morsmålsordbok som byggjer på vitskaplege samlingar. Jamstillingsvedtaket i Noreg 1885 mellom "Landsmaalet og vort almindelige Bogsprog" ville ha vore utenkjeleg om ikkje Ivar Aasens grammatikkar og ordbøker alt låg føre. Dei afrikanske språka i Zimbabwe ville ikkje ha fått jamstelt status i grunnlova frå 2013 utan ordboksarbeidet på 1990-talet (Chabata 2015:20).

Dei store morsmålsordbøkene (a) viser semje i språksamfunnet om samanheng mellom ordform og tyding, (b) viser fram skriftspråket i bruk, (c) populariserer språkvitskapleg terminologi på morsmålet, (d) legg eit sakleg grunnlag for morsmålsopplæring, (e) gjer det mogleg å setja krav til målbruk i det offentlege rommet.

Føresetnaden for at den store morsmålsordboka skal ha desse funksjonane, er at det som står i ordboka, byggjer på faktisk språkbruk og sams normer. Morsmålsbrukarar slår sjeldan opp for å kontrollera tydinga av orda dei bruker. Dei veit kva orda tyder, gjennom ein heil oppvekst og mange års utdanning med morsmålstrening. Men når det oppstår usemje, blir den store morsmålsordboka oppmann.

10. Verdi for forskning og språknormering

Målet for dei vitskaplege ordboksverka i Norden er å gje tilnærma heildekkjande framstillingar av kvar sine språk som produkt av ei lang utviklingshistorie. Finst det no ei heildekkjande framstilling av ordtilfanget i norsk? I hovudsak må svaret bli ja, men to viktige arbeidsfelt står att. Det eine er det at alfabetbolken a-h i NO ikkje er tilpassa det digitale formatet. Teksten må gjennomarbeidast med dette målet for auga, og i samband med ei slik gjennomarbeiding er det naturleg å revidera innhaldet. Dei eldste

delane av NO vart redigerte for 50-60 år sidan på eit veikare underlagstilfang enn det som no ligg føre. Også i dei nyare delane vil systematikken vera mindre streng enn i den delen av ordboka som er redigert direkte inn i databaseformatet. Det andre store tomrommet er at det ordtilfanget som er særmerkt for bokmål, per dato ikkje er tilgjengeleg i ei vitenskapleg ordbok som fyller dagens krav.

Kva verdi har den vitenskaplege leksikografien for vidare forskning? Vitenskaplege ordboksverk er tilrettelagde framstillingar av språket på mange plan, og har i alle år vore viktige hjelpemiddel i forskinga, som syntesar av føreliggjande informasjon om einskildord og ordgrupper. Skiljet mellom før og no ligg i at ein kan ein dra ut alle kategorikombinasjonar frå ein leksikalsk database, og få ut store materialmengder sortert og ordna på nokre sekund. Det er lett å få tak i bearbeidd inventar av alle slags ordgrupper i norsk. Substantiv med fleire genus? Prefiks på veg over til å bli samansetningsledd (av typen *trans-*)? Er verb på *-na* som hovudregel gamle inkoativ? Synonym for *kvinne*? Ein leksikalsk database leverer raske svar, eller råmateriale for svar, på slike spørsmål. Ein kan vera usamd i svaret, men ein får eit startpunkt.

Det er fleire tiår sidan siste freistnad vart gjort på beskriving av ordlagingsmekanismane i norsk. Bøkene var solide verk for si tid, men dei vart skrivne utan tilgjenge til leksikalske databasar eller korpus. For studiar i fraseologi, semantikk, ordnett og omgrepshierarki er likeins ein leksikalsk database eit godt utgangspunkt.

11. Eigenverd for språksamlingane

I ei ideell verd burde den vordande omlegginga til IKT alt på 80-talet ha ført til ein prinsippdiskusjon om tilhøvet mellom ordboksprosjekta og morsmåls-samlingane. Det burde ha vore mogleg å innsjå alt då at språksamlingar, og den kompetansen som er bygt opp i tilknytning til dei, har eigenverd og fleirfunksjonalitet for vedlikehald og oppfølging av standardspråka i dei aktuelle språksamfunna, uavhengig av aktuelle ordboksprosjekt. Fram til om lag 1980 vart deponering av innsamla talemålstilfang praktisert ved Universitetet i Oslo – svært mykje av målføreregistreringa er gjort av hovudfagsstudantar – men innsikta var ikkje formulert som program i dei

aktuelle fagmiljøa. No er det som nemnt kome krav frå Forskingsrådet at alt innsamla tilfang skal takast vare på i minst ti år. Dette er alt for lite. Språksamlingane som ligg til grunn for dei store ordboksverka har ofte meir enn 100 år attom seg, og krev same tidsperspektiv i framtida som andre sentrale arkivkomponentar, som til dømes riksarkiv og statistiske sentralbyrå. Liksom desse institusjonane høyrer språksamlingane til den sentrale infrastrukturen som må vera på plass og i funksjon for at samfunnet skal kunna fungera.

Samstundes må språksamlingane haldast levande, gjennom prosjektutvikling, undervisning og individuell forskning. Målførearkivet ved Universitetet i Oslo – no nedlagt – tilbød arbeidsmiljø og rettleiing for meir enn 100 hovudfagsstudentar. Anten det var fonologi, morfologi, leksikon eller saksorientert språkinnsamling som var studentens perspektiv og ramme, så hadde arbeidsinnsatsen avkastning i gjenbruk. Tilfanget som studentane la til grunn for eige arbeid, vart integrert i språksamlingane og er nytta om att mange gonger, mellom anna i NO. Sunn fornuft tilseier at språksamlingane igjen blir depot for innsamla språkdata, og til gjengjeld tilbyr bistand med å gjera datainnsamling mest mogleg funksjonell, både for feltarbeidaren og for den større forskarverda.

12. Avrunding

Morsmåla er under press i mange land i Europa, ikkje minst i dei nordiske landa. Domenetap er ein reell fare, for ikkje å sei ein pågåande prosess, og kanskje særleg ved universiteta. Det er derfor umogleg å sjå bort frå den språkpolitiske funksjonen som dei vitskaplege ordboksverka har berre ved å liggja føre, anten dei tek for seg standarspråk eller målføre. Digitalisering skaper ein kontakt med brukarane som ikkje fanst før, og gjer alle påstandar om irrelevans og manglande legitimitet til skamme.

Framfor alt syner ordboksverka fram morsmål og talemål som reiskapar for tanken. Eit språksamfunn kan ikkje skifta språk utan å seia farvel til si eiga kognitive historie, og det vil folk ikkje. Det er for dumt. Eg er derfor optimist på vegner av den vitskaplege morsmålsleksikografien, i seg sjølv og som grunnlag for privat og kommersiell utvikling av språkprodukt. I framtida vil vonleg sambandet mellom språksamlingar, ordboksverk,

forskning og utdanning bli styrkt, til beste for kunnskaps- og omgrepsutvikling på nordiske språk i den nordiske landa.

Litteratur

Ordbøker

- Cawdrey, Roger (1604): *The first English Dictionary*. Reprint 2007. Introduction by John Simpson, Chief Editor Oxford English Dictionary. Oxford: The Bodleian Library.
- DAF (1694-) = *Dictionnaire Academie Francaise*. <www.academie-francaise.fr/le-dictionnaire/la-9e-edition>.
- DWB (1854-1961) = *Deutsches Wörterbuch*. Red. J. Grimm, W. Grimm. Leipzig: Hirzel.
- Elyot, Sir Thomas (1538): *Dictionary* [Latin-English Dictionary]. London: printed by Thomas Berthelet.
- Jenssøn, Christen (1646): *Den Norske Dictionarium eller Glosebog*. Ny utgåve ved Torleiv Hannaas 1915. Kristiania (Oslo): Den norske historiske kildekriftkommission / Grøndahl & Søn.
- Johnson, Samuel (1755): *A Dictionary of the English Language*. London.
- Maaler, Josua (1561): *Die Teütsch spraach: alle wörter, namen, und arten zuo reden in Hochteütscher spraach: dem ABC nach ordenlich gestellt, unnd mit guotem Latein gantz fleissig unnd eigentlich vertolmetscht, dergleychen bishär nie gesähen. durch Josua Maaler*. Tiguri. Preface by Conrad Gesner. Reprint 1971. Hildesheim: Olms.
- NO (1950-2016) = *Norsk Ordbok. Ordbok over det norske folkemålet og det nynorske skriftmålet*. I – XII. Hovudredaktørar: A. Hellevik, L. S. Vikør, O. Grønvik, L. Killingbergtrø, D. Worren, H. Gundersen. Oslo: Det norske Samlaget.
- ODS (1918-1956) = *Ordbog over det danske Sprog*. Udgivet af Det Danske Sprog- og Litteraturselskab. København: Gyldendal.
- ODS-S (1992-2005) = *Ordbog over det danske Sprog. Supplement 1-5*. Udgivet af Det Danske Sprog- og Litteraturselskab. København: Gyldendal.

- OED (1884 -1928) = *Oxford English Dictionary*. Chief editors Murray, J. A. H. et al. Oxford: Oxford University Press.
- SAOB (1898-) = *Svenska Akademiens Ordbok över Svenska Språket* [37 av 39 band]. Utgiven av Svenska Akademien. Glerup.
- VDC (1611-1923) = *Il Vocabolario degli Accademici della Crusca*. Edizione elettronica. <vocabolario.sns.it/html/_s_index2.html>.
- WNT (1864-1998) = *Woordenboek der Nederlandsche Taal*. Instituut voor Nederlandse Lexicologie.
- Aasen, Ivar (1873): *Norsk Ordbog med dansk Forklaring*. Fjerde uforandrede Udgave 1918. Vestmannalaget. Kristiania: Cammer-meyers Forlag – Lars Swanstrøm.

Annan litteratur

- Arens, Hans (1984): *Aristotle's Theory of Language and its Tradition*. Studies in the History of the Language Sciences 29. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Arens, Hans (2000): Sprache und Denken bei Aristotles. I: *History of the Language Sciences. An international Handbook on the Evolution of the Study of Language from the Beginnings to the Present*. Vol. I. Berlin – New York: Walter de Gruyter, 367-375.
- Bakken, Kristin (2006): Relevans, legitimitet og resultatkrav – eit komparativt blick på utfordringane for dei store nasjonale ordboksverka i dag. I: Henrik Lorentzen & Lars Trap-Jensen (red.): *Nordiske studier i leksikografi 8. Rapport fra Konference om leksikografi i Norden, Sønderborg 24.-28. maj 2005*. Skrifter udgivet af Nordisk Forening for Leksikografi. Skrift nr. 9, 11-24.
- Brewer, Charlotte (2006): *OED Sources*. I: Lynda Mugglestone (red.): *Lexicography and the OED. Pioneers in the Untrodden Forrest*. Oxford: Oxford University Press, 40-58.
- Bussmann, H., G. Trauth & K. Kazzazi (2006): *The Routledge Dictionary of Language and Linguistics*. London and New York: Routledge.
- Bywater, Ingram (1962): *Artistotle on the Art of Poetry*. With a preface by Gilbert Murray. Oxford: The Clarendon Press.
- Bø, Reidar. (1989): Arbeidet med Norsk Ordbok. I: *Ord og mål. Festskrift til Magne Rommetveit*. Oslo: Kringkastingsringen, 80-94.

- Chabata, Emmanuel (2015): Current lexicographic trends: Achievements and challenges for the indigenous languages of Zimbabwe. I: Jan Ragnar Hagland & Åse Wetås (red.): *Ivar Aasen ute og heime - om moderne språkdokumentasjon etter Ivar Aasen*. Trondheim: Det Kongelige Norske Videnskabers Selskab, 7-27.
- Fjeld, Ruth Vatvedt (2008): Er norsk et fattig språk? I: *Utdanningsnytt* 20.05.2008. <utdanningsnytt.no/debatt/2008/mai/er-norsk-et-fattig-sprak/>.
- Grønvik, Oddrun (1998): Eit trehundreårsminne: Den første freistnaden på systematisk ordinnsamling i Noreg. I: *Ord om ord* 4. Årsskrift for leksikografi, 22-28.
- Grønvik, Oddrun (2005a): Redigeringsprogrammet for Norsk Ordbok 2014. I: Ruth Vatvedt Fjeld & Dagfinn Worren, (red.): *Nordiske studiar i leksikografi 7. Rapport frå Konferanse om leksikografi i Norden, Volda 20.-24. mai 2003*. Skrifter utgjevne av Nordisk foreining for leksikografi. Skrift nr 8, 156-165.
- Grønvik, Oddrun (2005b): Norsk Ordbok 2014 from manuscript to database - Standard Gains and Growing Pains. I: F. Kiefer, G. Kiss & G. Pajzs (eds.): *Papers in Computational Lexicography - Complex 2005*. Budapest: Linguistics Institute, Hungarian Academy of Science.
- Grønvik, Oddrun (2010): Leksikaliserte ordsamband i norsk. I: Harry Lønnoth & Kristina Nikula (red.): *Nordiska studier i leksikografi 10. Rapport från Konferensen om leksikografi i Norden, Göteborg 3-5 juni 2009*. Skrifter utgjevna av Nordiska föreningen för leksikografi. Skrift nr. 11, 118-129.
- Grønvik, Oddrun & Helge Gundersen (2015): Redigeringshandbok for Norsk Ordbok 2014. <no2014.uio.no/eNo/tekst/redigeringshandboka/redigeringshandboka.pdf>.
- Grønvik, Oddrun & Christian-Emil Smith Ore (2016): Nyskapinga i elektroniske ordbøker og språksamlingar. I: Helene Urdland Karlsen, Lars S. Vikør & Åse Wetås (red.): *Livet er æve, og evig er ordet. Festskrift til Norsk Ordbok*. Oslo: Det Norske Samlaget, 132-141.
- Hanks, Patrick (2010): Lexicography, Printing Technology, and the Spread of Renaissance Culture. I: Anne Dykstra & Tanneke Schoonheim (eds.): *Proceedings of the 14 International Euralex Congress*. Leeuwarden: Frisian Institute, 988-1006.

- Hellevik, Alf (1956): På skattegraving i vårt eige mål. I: *Syn og Segn*, 337-352. [Også utgjeve som særprent].
- Historisk-filosofisk fakultet, Utvalget for forskings spørsmål (1989): Vurdering av ordbok- og arkivarbeid i forhold til vitenskapelige kvalifikasjoner. 20.2.1989. Uttalelse. I manuskript.
- Karlsen, Helene Urdland, Lars S. Vikør & Åse Wetås (2016): *Livet er æve, og evig er ordet. Festskrift til Norsk Ordbok*. Oslo: Det norske Samlaget.
- Landau, Sidney (2001): *The Art and Craft of lexicography*. 2nd ed. Cambridge: Cambridge University Press.
- Mooijaart, Marijke (2013): A history of Dutch lexicography. I: *Trefwoord, tijdschrift voor lexicografie* 1, 1-34. <www.frysk-akademy.nl/trefwoord>.
- Mugglestone, Lynda (2006): 'Pioneers in the untrodden forrest': The *New English Dictionary*. I: Lynda Mugglestone (red.): *Lexicography and the OED. Pioneers in the Untrodden Forrest*. Oxford: Oxford University Press, 1-21.
- Murray, K. M. E. (2001; 1. ed. 1977): *Caught in the web of words. James Murray and the Oxford English Dictionary*. New Haven – London: Yale University Press.
- Noregs forskingsråd (2014): *Tilgjengeliggjøring av forskningsdata*. Policy for Norges forskingsråd. Oslo.
- Ore, Christian-Emil Smith & Nina Kristiansen (red.) (1998): *Dokumentasjonsprosjektet. Sluttrapport*. Oslo: Universitetet i Oslo.
- Ore, Christian-Emil Smith (2000): Metaordboka. I: *Ord om Ord*. Seksjon for leksikografi og målføregransking, 30-32.
- Ore, Christian-Emil Smith (2001): Metaordboken – et rammeverk for Norsk Ordbok? I: Martin Gällerstam, Kristin Johannesson, Bo Ralph & Lena Rogström (red.): *Nordiska Studier i Lexikografi 5, Rapport från Konferens om lexicografi i Norden Göteborg 27.-29. maj 1999*. Skrifter utgivna av Nordiska föreningen för lexicografi 6, 250-270.
- Osselton, Noel (2006): Murray and his European Counterparts. I: Lynda Mugglestone (red.): *Lexicography and the OED. Pioneers in the Untrodden Forrest*. Oxford. Oxford University Press, 59-77.
- Palmer, D. J. (1965): *The rise of English Studies*. London: Oxford University Press.
- Simpson. John (2007): Introduction. I: Cawdrey (1604), 8-31.

- Skolt, Knut & Margit Halbjørhus Wøllo (2013): *Gamle ord og uttrykk frå Hemsedal. Alfabetisk ordliste med lyduttale*. (1. utgåve 1999.) 2. utgåve. Hemsedal: Hemsedal Mållag.
- Steinfeld, Torill (1986): *På skriftens vilkår. Et bidrag til morsmåls-fagets historie*. Oslo: Landslaget for norskundervisning – Cappelen.
- Vikør, Lars (1999): Fleirgenerasjonsordbøker og tida. I: Peter Slotte, Pia Westerberg & Eva Otava (red.): *Nordiska Studier i Lexikografi 4. Rapport frå Konferensen om lexikografi i Norden, Esbo 21.-24. maj 1997*. Skrifter utgivna av Nordiska föreningen för lexikografi 5, 395-406.
- Vikør, Lars (2016): Språkdokumentasjon og språkforskning. To sider av same sak, eller to saker? I: Helene Urdland Karlsen, Lars S. Vikør & Åse Wetås (red.): *Livet er æve, og evig er ordet. Festskrift til Norsk Ordbok*. Oslo: Det Norske Samlaget, 105-131.
- Vikør, Lars & Åse Wetås (2016): Historia om Norsk Ordbok: om folket – av folket – for folket. Historia om ordboksverket fram til i dag. I: Helene Urdland Karlsen, Lars S. Vikør & Åse Wetås (red.): *Livet er æve, og evig er ordet. Festskrift til Norsk Ordbok*. Oslo: Det Norske Samlaget, 15-39.

Oddrun Grønvik
seniorforskar
Universitetet i Oslo
Postboks 1102 Blindern
NO-0315 Oslo
Oddrun.gronvik@iln.uio.no

TEMA

DEN DIGITALE
ORDBOGSBRUGER

Fordele og ulemper ved løbende opdatering af den digitale Retskrivningsordbog

Margrethe Heidemann Andersen, Jørgen Nørby Jensen & Anita Ågerup Jervelund

Once a year, the digital version of Retskrivningsordbogen (the official dictionary of Danish Orthography) is updated with at least 150 new entries. The orthographical rules and the spelling of the existing entries in the dictionary remain unaltered. However, imprecisions in the orthographical system caused by the updating may occur. In this article we describe how we update the dictionary, and we discuss the problems these updates may cause.

1. Indledning

Fra og med 2014 bliver den digitale udgave af Retskrivningsordbogen (RO) en gang om året opdateret med mindst 150 nye opslagsord. Det er dog ikke helt uproblematisk at udvide den officielle Retskrivningsordbog løbende. Selvom der i princippet ikke ændres i de eksisterende former, bliver den gældende norm alligevel løbende udfordret i kraft af tilføjelsen af nye ord og nye eksempler, og i nogle tilfælde opstår der skævheder i normen.

I denne artikel vil vi beskrive hvordan vi opdaterer RO løbende, og vi vil give eksempler på hvad en sådan årlig opdatering gør ved den officielle retskrivningsnorm, som skolen og den offentlige sektor er forpligtede til at følge.

2. Hvordan opdateres RO?

Efter at den seneste Retskrivningsordbog var udkommet i 2012 i både trykt og digital form, begyndte Sprognævnet at drøfte hvordan den løbende opdatering af ordbogen skulle foregå. Der blev udarbejdet en strategi for opdateringen, og denne strategi blev drøftet på både institutmøder, bestyrelsesmøder og fagrådsmøder i Sprognævnet inden den blev vedtaget i

sensommeren 2013. Den første opdatering af RO blev offentliggjort i december 2014.

Ifølge strategien skal den digitale udgave af RO opdateres én gang om året i december måned ved at

- der indsættes nye opslagsord, ca. 150 ord om året
- der tilføjes oplysninger i eksisterende artikler, fx sammensætnings- og brugseksempler, betydninger samt bøjningsformer hvor de mangler
- der tilføjes eksempler e.l. i en retskrivningsregel for at tydeliggøre reglen.

Det fremgår tydeligt af strategien at den årlige opdatering ikke må indeholde normændringer, så der er altså ingen ændringer i staveformer, bøjningsformer, sammensætningsformer eller retskrivningsreglerne.

Af hensyn til især skoler og hele den offentlige sektor, som er forpligtede til at følge den officielle retskrivning som fastsat i RO, indføres der ikke normændringer løbende. Det er ikke rimeligt hvis sprogbrugerne skulle sætte sig ind i den slags retskrivningsændringer en gang om året. Det må vente ca. 10 år til der udkommer en helt ny RO i ca. 2022, som påpeget i strategien.

I strategien er det nærmere beskrevet hvilke tilføjelser der i givet fald kunne betragtes som normændringer, og hvilke der ikke kan, for det er ikke indlysende hvor grænsen går. I det følgende vil vi se nærmere på disse tilladte og ikketilladte tilføjelser ifølge strategien.

3. Tilføjelse af nye opslagsord

Det primære formål med den årlige opdatering er at få nye opslagsord i ordbogen. De fleste nye opslagsord følger et i forvejen normeret mønster og indgår fint i de almindelige retskrivningsregler. Det gælder fx *chokoladebar*, *e-cigaret* og *genforskning*, som kom med i opdateringen i 2014. Ordene *chokoladebar* og *genforskning* er sammensætninger hvor hvert led findes som simpleksord i RO i forvejen og således har en normeret stavemåde. Ordet *e-cigaret* følger andre ord med *e-* som førsteled, fx *e-bog* og *e-handel*, der alle skrives med bindestreg pga. forkortelsen *e*. Det er uproblematisk.

Flere af de nye opslagsord stammer ikke overraskende fra engelsk, og nogle af disse ord kræver lidt flere undersøgelser end de mere hjemlige ord.

Det gælder et ord som *tweete/twitte*, der endte med at blive opført som en dobbeltform fordi vores undersøgelser i forbindelse med opdateringen i 2014 viste at begge former er almindelige. Opslagsord der har en dobbeltform, kan altså også uden videre indsættes i ordbogen. Men hvis det viser sig at den ene dobbeltform forsvinder fra den faktiske sprogbrug, hvilket muligvis er ved at ske med formen *twitte*, kan vi ikke fjerne den fra ordbogen før der udkommer en ny udgave i ca. 2022.

4. Tilføjelse til eksisterende artikler

Det er muligt at indføre en del forskellige typer opdateringer i de eksisterende artikler. Nogle af disse opdateringer, fx tilføjelse (sletning el. ændring) af betydningsoplysninger, har ingen indflydelse på den ortografiske norm, og derfor forpligter de ikke på samme måde som opslagsformer, bøjningsformer og sammensætningsformer.

Der er især behov for at tilføje bøjningsformer og sammensætningsformer løbende, og strategien nævner specifikt at der kan tilføjes en bøjningsform dér hvor den mangler (eller er glemt). Typisk vil det dreje sig om tilføjelser af flertalsformer:

deficit *sb.*, -tet (FORMELT *underskud*) >

deficit *sb.* -tet, -ter *el.* deficit, -te(r)ne (FORMELT *underskud*)

Men det kan også være en tilføjelse af bestemt form singularis hvis det er den oplysning der mangler, jf. formen *itemmet* under opslagsordet *item*.

På samme måde som ved bøjningsformer kan der tilføjes sammensætningsformer hvis der ikke findes oplysninger om disse allerede:

kyllingekød *sb.*, -et >

kyllingekød *sb.*, -et, *i sms.* kyllingekød- *el.* kyllingekøds-, *fx*

kyllingekød(s)ret

Se Jervelund (2014) for yderligere eksempler på tilladte tilføjelser.

5. Ikketilladte tilføjelser og opdateringer

De tilføjelser som bliver betragtet som normændringer, er ændringer af eksisterende former eller tilføjelser til eksisterende former. Det gælder såvel opslagsords staveformer som bøjningsformer og sammensætningsformer. Det er herudover heller ikke muligt at indføre ændringer i retskrivningsreglerne.

Ved den årlige opdatering er det altså ikke muligt at gå fra en staveform til en anden. Hvis vi ønskede at ændre stavemåden *lav* 'sammenslutning' til *laug*, som mange sprogbrugere foretrækker – til trods for at den følger nogle meget forældede retskrivningsprincipper og ikke på noget tidspunkt i historien har været at finde i en officiel retskrivningsordbog – ville det ikke kunne lade sig gøre foreløbig.

Tilføjelser til eksisterende former er som sagt heller ikke tilladte ifølge strategien. Vi ville derfor ikke kunne indføre fx *tidsløs* som opslagsform ved siden af *tidløs*, selvom *tidsløs* er forholdsvis almindelig i den faktiske sprogbrug (jf. også dobbeltformerne *tidnød/tidsnød*, *tidmangel/tidsmangel*, *tidrøver/tidsrøver* m.fl.). En tilføjelse af en dobbeltform er altså ikke mulig, ligesom en sletning af en dobbeltform i øvrigt heller ikke er det.

Hvis et ord allerede er angivet med en flertalsform, kan vi selvsagt ikke udskifte denne form med en anden, men vi kan heller ikke tilføje en ny flertalsform ved siden af den eksisterende form. Vi kan fx ikke indføre flertalsformen *shops* ved siden af den eksisterende *shopper* (jf. afsnit 6.2).

Alt hvad der drejer sig om dobbeltformer i bred forstand, dvs. to side-stillede opslagsformer, bøjningsformer eller sammensætningsformer, kan der med andre ord ikke røres ved i en opdatering. Nye opslagsord kan dog godt indføres som dobbeltformer (jf. eksemplet *tweete/twitte* ovenfor). Man kan således opsummerende sige at en RO-opdatering må gå fra unormeret eller fra "en tom plads" til normeret eller "en udfyldt plads", men ikke fra én normering til en anden.

6. Inkonsekvenser i retskrivningsnormen

Som nævnt kan der uden videre indsættes nye opslagsord i den digitale udgave af RO. Principielt skulle der være tale om en forholdsvis uproblematisk opgave, men det er der ikke nødvendigvis.

Blandt de 212 nye opslagsord der blev indført i RO ved opdateringen i december 2014, finder man eksempelvis substantivet *hanggliding*. Ordet må siges at tilhøre rigssprogets almindelige ordforråd, og det er derfor rimeligt at det kan slås op i RO. Skrivemåden er for en umiddelbar betragtning uproblematisk. Den svarer til den almindelige (engelsknære) udtale af ordet, jf. fx udtaleoplysningerne i Den Danske Ordbog, og det er også den stavemåde der bruges i praksis.

Siden 1996 har man imidlertid kunnet slå ordet *paraglidning* op i RO. Dette ord har en dansk afledningsendelse (-*ning*), og herved bliver to ord der både etymologisk og orddannelsesmæssigt ligner hinanden, behandlet forskelligt.

Til grund for stavemåden *paraglidning* ligger en dansk udtale af ordet: ['pa:a.gliðneŋ]. Denne udtale er nævnt i Den Danske Ordbog, men det tilføjes at ordet også forekommer i formen *paragliding* med udtalen ['para.gljɔjden]. Og denne form er langt den hyppigste i den faktiske sprogbrug. En søgning i avisdatabasen Infomedia med periodeafgrænsningen 1.1.1990-17.6.2015 gav således i alt 1584 eksempler på den ukorrekte skrivemåde *paragliding*, mens der blot var 115 eksempler på den korrekte *paraglidning*. På den baggrund ville det ikke være urimeligt at ændre *paraglidning* til *paragliding* i RO, eller i det mindste tillade formen *paragliding* ved siden af *paraglidning*. Hvis man ændrer én staveform til en anden, vil der imidlertid være tale om en normændring, og i den løbende opdatering af ordbogen kan der ikke foretages normændringer. Det samme gælder tilføjelse af en dobbeltform (jf. afsnit 5). En eventuel ændring må således vente til der kommer en ny udgave af ordbogen. Indtil da må brugerne af RO leve med at to beslægtede ord rent retskrivningsmæssigt opfører sig forskelligt, og man kan hævde at der herved er opstået en skævhed i retskrivningsnormen.

6.1. Baggrunden for formen *paraglidning*

Hvorfor *paraglidning* i sin tid blev indført som eneform i RO, fortæber sig lidt i det uvisse, men det hænger formodentlig sammen med at der i Sprognævnets ordsamling faktisk var flere citater med denne stavemåde. Ordsamlingen udgjorde især tidligere et vigtigt grundlag for fastlæggelsen af retskrivningen; nu suppleres der med en lang række andre kilder. En anden grund kan være de fordanskningsbestræbelser som igennem hele den officielle retskrivnings historie har spillet en større eller mindre rolle. I *Bekendtgørelse om Dansk Sprognævns virksomhed og sammensætning* kan man således læse ”at fremmedord, der er blevet almindelige i dansk, skrives i overensstemmelse med de regler, der gælder for oprindelige danske ord og ældre låne- og fremmedord i dansk. Dette princip gælder først og fremmest ord fra græsk, latin og fransk og ord dannet af græske, latinske og franske bestanddele. Ord fra andre sprog, specielt nyere ord, skrives som hovedregel i overensstemmelse med skrivemåden i det pågældende sprog eller med international praksis” (§ 1, stk. 3).

Engelske låneord (der her gemmer sig under betegnelsen ”ord fra andre sprog”) får således som hovedregel ikke en dansk stavemåde, men der sker alligevel en vis ortografisk tilpasning. Når et engelsk låneord har vundet indpas i sproget, tages der fx ikke hensyn til den engelske praksis med at særskrive betydningsmæssige enheder. Vi skriver fx *artdirector*, *clockradio*, *heavyrock*, *exitpoll*, *cheerleader* og *matchfixing*. Engelsk har *art director*, *clock radio*, *heavy rock*, *exit poll*, *cheer leader* og *match fixing*.

I forhold til afledning er billedet lidt mere broget, men der har tidligere været flere engelske låneord med dansk afledning i RO, fx *jogning* (ved siden af *jogging*) i RO 1986 og RO 1996 og *hackning* (ved siden af *hacking*) i RO 1996 og RO 2001. Ved visse ord er der i RO 2012 både en dansk og engelsk form, fx *booking/bookning*, *headhunting/headhunting*, og ved andre kun en engelsk, fx *downloading*. Den modsatte udvikling ses ved ordet *mobning*. Da ordet kom ind i dansk i begyndelsen af 1970’erne, kunne man også støde på formen *mobbing* (jf. Nye Ord i Dansk 1955 til i dag), men i dag er den danske form enerådende. Trods en vis variation på området er hovedtendensen dog klar: Der bliver færre fordanskede former og flere genuine. Også af disse grunde kunne en ændring af *paraglidning* til *paragliding* være på sin plads.

6.2. Bøjningen af nye opslagsord i RO

Også hvad angår bøjningen af nye opslagsord, kan der opstå problemer. Ved den kommende opdatering af RO i december 2015 vil vi efter al sandsynlighed indføre ordet *webshop*. Ordet skal skrives på den anførte måde, men hvad med bøjningen? Ved at se på hvordan et beslægtet ord som *workshop* er beskrevet i RO, kan man nå frem til at der i pluralis bør være valgfrihed mellem *webshopper* og *webshops*. Ifølge RO kan *workshop* nemlig hedde både *workshops* og *workshopper* i pluralis. Men hvis man i stedet lader ordet *shop* være mønster for bøjningen, kan man nå frem til at det kun kan hedde *webshopper* i pluralis. Ordet *shop* har været med i RO siden 1986 og har hele tiden kun haft pluralisformen *shopper*.

Ordet *workshop* har i øvrigt ikke altid haft mulighed for *-s* i pluralis i den officielle retskrivning. Det fik det først med RO 2012. Her blev der ved en lang række substantiver der stammer fra engelsk, indført pluralis-s'er ved siden af (eller i stedet for) danske pluralisbøjninger (jf. Nørby Jensen 2013), fx *jobs* (ved siden af *job*), *teams* (ved siden af *team*), *hooligans* (ved siden af *hooliganer*) og altså *workshops* (ved siden af *workshopper*).

Dansk Sprognævn har principielt ikke noget imod ord fra andre sprog, men nævnet har ikke desto mindre traditionelt anset det for ønskværdigt hvis sådanne ord i så høj grad som muligt kunne tilpasses grammatisk, dvs. mht. bøjning, til det danske sprogsystem. RO har derfor som hovedregel været temmelig tilbageholdende med at tillade pluralis-s'er. Som de ovenstående eksempler antyder, har ordbogen undertiden været så tilbageholdende at det næsten kunne betragtes som en form for morfologisk purisme. Med Henrik Galberg Jacobsens ord har det altid været "lidt pinligt at det ellers fornuftige og upuristiske Sprognævn ikke i højere grad anerkendte den faktiske sprogbrug på dette punkt" (Galberg Jacobsen 2013: 287). Pluralisformer som *jobs*, *teams*, *workshops* osv. har i hvert fald for længst vundet indpas i det almindelige skrift- og talesprog, også blandt gode og sikre sprogbrugere, og de blev derfor langt om længe anerkendt i RO 2012. Ordet *shop* blev imidlertid forbigået, og vi har nu den situation at ordene *shop* og *workshop* rent bøjningsmæssigt behandles forskelligt i RO. Sandsynligvis vil vi lade ordet *webshop* følge *workshop* og altså tillade både *webshops* og *webshopper* i pluralis. Derimod kan vi ikke indføre pluralisformen *shops*, for en tilføjelse til en eksisterende form er at betragte som en normændring (jf. afsnit 5).

7. Normændring eller ej?

Et yderligere problem er at det undertiden kan være svært at afgøre hvad der er en normændring, og hvad der ikke er. Lad os sige at vi ved opdateringen i december 2015 vil sætte ordet *tantralsex* ind i RO. Stavemåden giver sig selv, men kan ordet fx bøjes i bestemt form singularis (*tantralsexen*)? I den forbindelse er det naturligvis oplagt at se på hvordan andre ord på (-)sex er behandlet i ordbogen. RO 2012 indeholder i alt seks ord på (-)sex (*sex*, *købesex*, *gruppesex*, *analsex*, *oralsex* og *telefonsex*). Ved fem af de seks ord står der blot at de er fælleskøn, og der er ikke angivet nogen bestemt form singularis. Af *Vejledning i ordbogens brug* (RO 2012, s. 16) fremgår det at der ved substantiver er ”givet oplysning om bestemt form singularis, ubestemt form pluralis og bestemt form pluralis når disse kendes”. Endvidere hedder det: ”Ved ord der næppe forekommer i bestemt form singularis, er der i parentes givet oplysning om ordets køn” (s. 17). Ifølge ordbogen forekommer de fleste af de nævnte ord altså næppe i bestemt form (*sexen*, *analsexen*, *telefonsexen* osv.). Ved opslagsordet *gruppesex* er formen *gruppesexen* imidlertid eksplicit angivet, men sådan har det ikke altid været. I den nedenstående figur kan man se hvordan ord på (-)sex er blevet behandlet i RO siden 1986:

RO 1986	RO 1996	RO 2001	RO 2012
			analsex sb. (fk.)
gruppe sex	gruppe sex sb., -en	gruppesex sb., -en	gruppesex sb., -en
		købesex sb., en	købesex sb. (fk.)
			oralsex sb. (fk.)
sex, en	sex, en	sex sb., en	sex sb. (fk.)
			telefonsex sb. (fk.)

Figuren viser flere ting. For det første er der sket en forøgelse af antallet af ord på (-)sex. For det andet kan man se at *sex* og *gruppesex* i 1986 havde samme (manglende) bøjning. Den lodrette streg i RO 1986 og RO 1996 markerede at oplysninger om ordets bøjning skulle findes under det ord der

stod efter den lodrette streg, i dette tilfælde *sex*. Der var kun givet oplysninger om bøjning mv. hvis de afveg fra oplysningerne under sidsteleddet som opslagsord. Mange brugere havde dog svært ved at forstå hvad strengen egentlig betød, og derfor blev den fjernet igen i RO 2001 (se Jervelund 2001). I 1996 blev der indsat bøjningsoplysninger ved *gruppensex*, men ikke ved *sex*, og sådan har det været lige siden. De øvrige ord med *sex* som andetled har fulgt bøjningsmønstret for simpleksordet *sex*, dvs. ingen bøjning. Imidlertid er det svært at se hvorfor fx *gruppensex* og *købesex* skal behandles forskelligt i ordbogen. Hvorfor kan man skrive *gruppensex*, men ikke *købesex*? Mest rimeligt ville det således være at behandle de pågældende ord på samme måde. Én mulighed er at lade ordet *gruppensex* danne mønster for de andre ord. I så fald skal der eksplicit angives bestemt form singularis ved alle de nævnte ord (*sexen, analsexen, oralsex* osv.). Der vil være tale om tilføjelser til eksisterende artikler, og det er en tilladt opdatering (jf. afsnit 4). En anden mulighed er at lade de andre ord danne mønster for beskrivelsen, og så skal bøjningsformen ved *gruppensex* fjernes, men det ville næppe kunne betragtes som en tilladt opdatering og må derfor vente til en ny udgave af RO.

De ovenstående eksempler afslører et paradoks ved den løbende opdatering af RO: På den ene side kan vi indsætte nye opslagsord oftere og hurtigere end nogensinde før, men på den anden side kan vi ikke altid foretage de nødvendige justeringer i normen.

8. Nye opslagsord i RO. Er det en god ide?

Forud for den første opdatering af RO udformede Sprognævnet sammen med analyseinstituttet CEM Institute – Voxmeter en omnibusundersøgelse hvor 1002 personer blev stillet en række forskellige spørgsmål, bl.a. om deres brug af digitale og trykte opslagsværker om dansk sprog. Hensigten med omnibusundersøgelsen var fra Sprognævnets side at finde ud af i hvor høj grad RO anvendes og kendes af danskerne. Respondenterne blev bl.a. spurgt om deres holdning til at Retskrivningsordbogen bliver opdateret én gang om året, og det blev nævnt at ord som *gastropub*, *makeover* og *selfie* snart ville være at finde i RO. Langt de fleste af de adspurgte, nemlig 69 %, mente at det var en god ide, mens 16 % mente at det var en dårlig ide, og 15 %

svarede ”ved ikke”. Men spørgsmålet er hvad det er respondenterne har svaret på: Er det at RO bliver opdateret løbende? Er det at RO bliver opdateret én gang om året (og ikke fx to gange om året)? Eller er det snarere at RO bliver opdateret med ord af typen *gastropub*, *makeover* og *selfie*, altså nye eller nyere ord, ofte engelske lån? Det fremgår af respondenternes begrundelser at det størstedelen af dem forholder sig til, er at der kommer nye engelske ord ind i ordbogen. Blandt de negative respondenter er der således flere der giver udtryk for at opdateringen ikke drejer sig om danske ord, men om engelske, og at man ikke synes at engelske ord som *selfie* og *makeover* burde komme ind i en dansk ordbog (altså RO), men at de i stedet burde findes i en slangordbog. Omvendt er der en del af de positive respondenter der svarer at det er godt at ord som *selfie* og *makeover* kommer ind i RO fordi sproget udvikler sig, og vi bliver nødt til at følge med udviklingen.

Selvom undersøgelsen viser at respondenterne primært har taget stilling til optaget af engelske ord i ordbogen, må man konkludere at danskerne generelt synes at det er en god ide at RO løbende bliver opdateret. Og man kunne forestille sig at hvis spørgsmålet havde været formuleret anderledes (hvis der fx også havde været eksempler på nye opslagsord som *cirkelspark* og *elcykel*), ville der måske have været endnu flere positive tilkendegivelser.

9. Sammensætninger i RO's søgelog

RO's søgelog kan bruges til at finde oplysninger om hvilke opslagsord brugerne leder forgæves efter. Vi har her begrænset os til at undersøge unikke søgninger i søgeloggen på én bestemt dag, nemlig d. 27. april 2015. Stikprøven indeholder knap 300 forskellige søgninger, og vi har her koncentreret os om de søgninger der handler om manglende sammensætninger i ordbogen.

Der er i RO visse begrænsninger med hensyn til udvalget af ord. En af de væsentligste vedrører ”de sammensætninger hvis betydning umiddelbart fremgår af de enkelte leds betydninger, fx *morgenbad*, *morgencomplet*, *morgenkaffe*, *morgentog*. Sådanne sammensætninger er ikke medtaget, medmindre de er meget gængse, fx *morgenkaffe*, eller de kan volde problemer med hensyn til stavning, bøjning eller orddannelse, fx *morgencomplet*” (Vejledning i ordbogens brug s. 11). Den konklusion som

folk der ikke læser vejledningen (og dem er der formentlig mange af), drager når de forgæves leder efter helt almindelige sammensætninger som *kaffedrikning* og *kaffefilter*, er ofte at der så må være tale om sammensætninger som enten ikke er gangbare, eller som ikke er en del af ”det officielle danske ordforråd”, og det udgør med Erik Hansens ord ”et nationalpædagogisk problem” (Hansen 2000:12). En løsning på dette nationalpædagogiske problem kunne være at indlemme flere af de sammensætninger som søgeloggen viser at brugerne har ledt forgæves efter, i ordbogen. Det drejer sig fx om sammensætninger som *dialogmøde*, *kvalitetsvare*, *lægemiddelmarked* og *sociokulturel*. Ingen af disse sammensætninger findes i ordbogen, men det gør *dialogboks*, *kvalitetsarbejde* og *lægemiddelindustri* til gengæld – og så burde det være ligetil at se hvordan de manglende sammensætninger staves. *Sociokulturel* findes ikke mere som opslagsord i ordbogen, men er rykket ned som sammensætningseksempel under orddelen *socio*-. Selvom RO i dag indeholder langt flere sammensætningsoplysninger end tidligere (Heidemann Andersen, Nørby Jensen & Jervelund 2014), kan man på baggrund af søgeloggen og den generelle viden om at mange ikke læser vejledningen i RO, diskutere om der er nok sammensætninger i ordbogen, og om det er hensigtsmæssigt at sammensætningseksempler ikke findes på alfabetisk plads (jf. Tarp 2013). Til den førstnævnte kritik kan man indvende at man kan blive ved med at sætte sammensætninger ind i en ordbog uden nogensinde at have en udtømmende liste, og at RO i det mindste giver hjælp til hvordan almindelige sammensætninger dannes. Når det er sagt, skal det dog ikke være nogen hemmelighed at vi overvejer muligheden af at sætte flere sammensætningseksempler ind i den digitale version af ordbogen fordi der jo ikke er de samme begrænsninger her i forhold til ordbogens størrelse som ved den trykte ordbog. Den sidstnævnte kritik er berettiget i og med at det er u hensigtsmæssigt at brugerne ikke umiddelbart finder en sammensætning som *sociokulturel* (der jo rent faktisk findes i ordbogen) sådan som den digitale Retskrivningsordbogs søgefunktioner er nu.

10. Konklusion

Sprognævnet har meldt klart ud at den årlige opdatering af RO ikke indeholder ændringer i den eksisterende retskrivningsnorm. For at kunne leve op

til dette har det været nødvendigt at fastlægge nøje hvilke opdateringer der kan lade sig gøre uden at ændre i normen. Det har vi forsøgt at formulere i en strategi for den løbende opdatering af RO. Nu har vi gjort os de første erfaringer med opdateringen, og det er muligt at strategien må udvides og præciseres så de problemer vi har beskrevet her i artiklen, formindskes. Men spørgsmålet er hvor meget vi kan gøre med god samvittighed og stadig påstå at der ikke er tale om løbende normændringer. Hvor går grænsen mellem fejlrettelser, forglemmelser og udbedringer af utilsigtede ortografiske skævheder på den ene side og egentlige normændringer på den anden side? Andre ordbøger kan nemt ændre og rette løbende, men RO er forpligtet på en helt anden måde med sin status som den officielle retskrivningsordbog.

Litteratur

Ordbøger

- Den Danske Ordbog* (2003-2005). <www.ordnet.dk/ddo>
Nye Ord i Dansk 1955 til i dag. <www.dsn.dk/noid>
Retskrivningsordbogen (4. udg., 2012). <www.dsn.dk/ro>

Anden litteratur

- Galberg Jacobsen, Henrik (2013): 2012-retskrivningen – i historisk perspektiv. I: *Gode ord er bedre end guld. Festskrift til Henrik Jørgensen i anledning af 60-års dagen*, 273-298.
- Hansen, Erik (2000): Normering og Sprognævn. I: Holmen, Anne & Jørgensen, J.N. (red.): *Sprogs status i Danmark i år 2011. Københavnerstudier i tosprogethed* 32. Danmarks Pædagogiske Universitet, 7-20.
- Heidemann Andersen, Margrethe, Nørby Jensen, Jørgen & Jervelund, Anita Ågerup (2014): Kommentar til Sven Tarp: Retskrivningsordbog på godt og ondt. I: *LexicoNordica* 21, 195-213.
- Jervelund, Anita Ågerup (2001): Stregen strøget! I: *Nyt fra Sprognævnet* 2001/4, 6-8

- Jervelund, Anita Ågerup (2014): 100 nye ord om året – tilføjelse af nye ord i RO. I: *Nyt fra Sprognævnet* 2014/2, 3-7.
- Nørby Jensen, Jørgen (2013): Ændrede bøjningsformer i Retskrivningsordbogen 2012. I: *Nyt fra Sprognævnet* 2013/1, 18-19.
- Tarp, Sven (2013): Retskrivningsordbog på godt og ondt. I: *LexicoNordica* 20, 185-200.

Margrethe Heidemann Andersen
seniorforsker, ph.d.
Dansk Sprognævn
Worsaaesvej 19, 4. sal
DK-1972 Frederiksberg C
heidemann@dsn.dk

Jørgen Nørby Jensen
seniorkonsulent, cand.mag.
Dansk Sprognævn
Worsaaesvej 19, 4. sal
DK-1972 Frederiksberg C
jn.jensen@dsn.dk

Anita Ågerup Jervelund
seniorkonsulent, cand.mag.
Dansk Sprognævn
Worsaaesvej 19, 4. sal
DK-1972 Frederiksberg C
aagerup@dsn.dk

The Case for Normalization: Linking Lexicographic Resources for Icelandic

Kristín Bjarnadóttir

The topic of this paper is the linking of two major lexicographic resources on Icelandic, the Dictionary of Old Norse Prose (ONP) in Copenhagen, and the Written Language Archive (WLA) at the Árni Magnússon Institute for Icelandic Studies in Reykjavík. The resources would be linked by normalizing the headwords from the ONP to the standard used for Modern Icelandic spelling by the use of a spellchecker originally designed for use on Modern Icelandic. This is possible, as the major differences in word forms between the two sources are mostly a matter of changes in spelling, whereas the morphology of Icelandic has been relatively unchanged through the centuries. Future work would consist of normalizing the citation collections themselves, thus creating a valuable resource linking variant word forms to their headwords or lemmas.

1. Introduction

The two major, longstanding lexicographic projects on Icelandic are the Dictionary of Old Norse Prose (ONP) in Copenhagen, and the work on a historical dictionary of Icelandic, as manifested in the Written Language Archive (WLA, Ritmálssafn Orðabókar Háskólans), in Reykjavík.¹ The line of demarcation between the projects is 1540, i.e. the year of publication of the first printed book in Icelandic. The projects were started in 1939 and 1944 respectively, and they are independent of each other, both in scope and methods. The topic of this paper is the linking of the data from these two

¹ The Dictionary of Old Norse Prose (ONP), originally established by The Arnarnagnæan Commission in Copenhagen, now part of the Department of Nordic Research, University of Copenhagen (Ellert Þór Jóhannsson et al., this issue), and The Written Language Archive, originally at The Institute of Lexicography (Orðabók Háskólans), now part of The Árni Magnússon Institute for Icelandic Studies, Reykjavík. For further information, see the websites: onp.ku.dk & arnastofnun.is. Thanks are due to the staff at the ONP and to Jón Friðrik Daðason.

resources by the normalization of the data in the ONP to standardized Modern Icelandic, as used in the lemmatization in the WLA. The first stage is the linking of the lemmas in both sources described below, but future work would include normalization of the citations themselves by automatic methods, thus creating a combined resource useful both in lexicography and in research on and access to data on the language in general.

2. The sister projects, ONP and WLA

Both projects, the ONP and the WLA, are traditional archives of dictionary slips, collected through decades in the latter part of the 20th century and gradually made available to users online, the ONP as an edited dictionary, the WLA as a collection of citations. The headwords in the ONP are normalized in accordance with the guidelines of the classic manner of Old Icelandic/Old Norse dictionaries, such as Fritzner (1867) and Cleasby & Vigfússon (1874), i.e. to the so-called ‘standard Old Norse spelling’ (‘samræmd stafsetning forn’), traditionally used in editions of medieval Icelandic texts. Figure 1 shows a sample from an entry in the ONP.

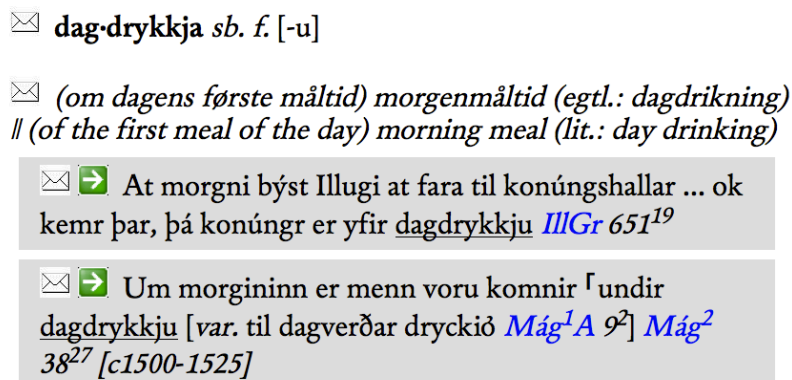


Figure 1: A sample entry from the ONP.

The headwords in the WLA are standardized to Modern Icelandic. For *dagdrykkja* the result is the same for the ONP and the WLA. Examples of differences in spelling are shown in Table 2.

Leitarorð: dagdrykkja		Raða eftir:	
Dæmi 1 - 1		undanfarandi	
af 1		orði	
Síða 1 af 1		eftirfarandi	
		orði	aldri
Nr	Dæmi	Orðmynd	Heimild
1	Prátt fyrir vanmetakennd og dagdrykkju er Guðný ekki búin að vera.	dagdrykkju	TímMM 1980, 254 Aldur: 20s
Dæmi 1 - 1 af 1 Til baka í "dagdrykkja" Ný leit			

Figure 2: A sample entry from the WLA.

The citations in both archives are shown with the original spelling from each source, ranging from editions of manuscripts from the 12th century to texts from the period 1980–1985, when traditional excerpting at the Institute of Lexicography was stopped and the focus shifted to the collection of computerized texts. The word forms of the headwords are clearly distinguished in the citations as seen in the examples from the websites in Figures 1 and 2, and links to the sources are given, making both the ONP and the WLA invaluable resources of lemmatized word forms. This data is a valuable language resource in itself, as it can be used to improve NLP tools by linking word forms irrespective of spelling.

3. The vocabulary of the ONP and the WLA

The respective number of headwords is 65.000 for the ONP, and over 600.000 for the WLA. Table 1 shows the proportions of headwords beginning in D in the automatic matching of the ONP and the WLA:

Headwords in WLA, total	16,815
Headwords in ONP, total	1,528
Headwords in ONP & WLA	853
Headwords in ONP only	675
Headwords in WLA only	15,962

Table 1: Figures for headwords beginning in D.

Changes due to discrepancies in lemmatization between the ONP and the WLA are not included in the figures in Table 1. These are due to different practices in the two projects, as in the creation of artificial headwords. This is the case when singular forms of nouns are used as headwords for nouns that only appear in the plural in (e.g., ONP *órr* masc. sg. vs. WLA *órar* masc. pl. ‘fantasy, delusion’), and when verbs are lemmatized as infinitives (active voice) in spite of only appearing in the mediopassive and/or as past participles. This method is traditional in older Icelandic dictionaries and it is more common in the ONP than in the WLA where it is rare. Discrepancies in the analysis of word class, such as adjectives vs. past or present participles of verbs, can also result in mismatches between the ONP and the WLA. Differences in word formation can also lead to mismatches, as in *dagaljós* and *dagsljós*, where the first constituent of the compound is either in the genitive singular or plural, without change of meaning. Table 2 shows differences in spelling and word formation in headwords, and one instance of a difference in meaning, i.e., in *dagdrykkja*.

ONP	WLA	Translation (ONP)	Notes
<i>dáðalauss</i>	<i>dáðalaus</i>	‘amoral’	cf. <i>dáðlaus</i>
<i>dáðlauss</i>	<i>dáðlaus</i>	‘without virtue ...’	cf. <i>dáðalaus</i>
<i>dáendi</i>	<i>dáindi</i>	‘fascination ..., miracle’	
<i>daga:fjöldi</i>	<i>dagafjöldi</i>	‘number of days’	
<i>dagaljós</i>	[→ <i>dagsljós</i>]	‘daylight’	cf. WLA: <i>dagsljós</i>
<i>dagsljós</i>	<i>dagsljós</i>	‘daylight’	
<i>dagdrykkja</i>	<i>dagdrykkja</i>	ONP: ‘morning meal’ WLA: ‘drink during the day’	

Table 2: Examples of headwords from ONP and WLA.

By resolving such discrepancies, the number of headwords from the ONP not found in the WLA can be reduced to a degree, but this work has to be done manually and it is quite time-consuming.

As the vocabulary of the ONP is not particularly large, it would be quite feasible to normalize and link the headwords manually, but the process can be made considerably faster by using automatic methods. This process can also serve as a part of the much more challenging project of normalizing running text from all periods of Icelandic language history.

The WLA was from the outset meant to be a historical dictionary for Icelanders and others studying the language from the 16th century onwards, and accordingly the headwords were normalized to Modern Icelandic. The headwords in the ONP are also normalized, but to a different standard of ‘normalized Old Icelandic’.

4. The cohesion of Icelandic morphology

The actual point of demarcation between the ONP and the WLA is unwarranted as far as language history is concerned, being set at the date of the first printed book in Icelandic, Oddur Gottskálksson’s translation of the New Testament (AD 1540), which is arbitrary from the point of linguistic changes in Icelandic.

The morphology of Icelandic has remained relatively stable through the centuries, both word formation and inflection, but the changes in spelling are extensive enough to make older texts very ‘difficult’ for modern day Icelanders, without training. The changes in spelling are regular (‘predictable’) from old to new, but not vice versa. With normalized modern spelling, older word forms are for the most part easily manageable for modern readers. In the first ten lines of the oldest Icelandic manuscript *Reykjaholtsmáldagi* (ca. AD 1150), 22 % of the word forms are in fact identical to the modern word forms, but the figure rises to 78 % by the use of 4 sets of simple rules, such as changing *o/v* to *u* in endings, changing the ending *-r* to *-ur*, and changing *þ* to *ð* except when word-initial (cf. Figure 3).

Original:

Til kirkio ligr irækiaholte heima land með ollo` landf nýtiom
þar fylgia kýr tottogo·griþungr tvevetr· xxx·a·oc hundraþ·

Modern character set:

Til kirkio ligr iravkiaholte heima land með ollom lands nytiom
þar fylgia kýr tottogo griþungr tvevetr xxx a oc hundraþ.

Modern spelling:

Til kirkju liggur í Reykjaholti <heimaland> með öllum <landsnytjum>.
Þar fylgja kýr tuttugu, griðungur tvævetur, 30 áa og hundrað.

Translation:

To the church in Reykjaholt belongs the home farm with all its produce.
 Along with that go twenty cows, a two-year old bull, and one hundred and
 thirty ewes.

Figure 3: The first two lines from Reykjaholtsmáldagi (AD 1250).

The spellchecker Skrambi described below makes this work feasible.²

4.1 Changes in spelling

The historical changes in Icelandic spelling are quite extensive, but here a small demonstration of the variants appearing in the entries of the word *drykkja* in the ONP and the WLA must suffice (cf. *dagdrykkja* in Figures 1 and 2 above).

- ONP: *dryckiom, drvckio, drvckio, dryckiar-, drykkior, dryckio, dryccior, dryckiu, dryckiona, drykkivr, drykkju, ðrykkiu, dryckiv ...*
 - [ck|k|kk|cc] → kk
 - i → j; v → u (in endings)
 - [u|o|v] → u
 - ð → d
- WLA: *dryckia, dryckiu, drykkja, drykkju ...*
 - ck → kk
 - i → j

Figure 4: Spelling variants in *drykkja* in ONP and WLA.

The changes in the ONP are much more extensive than in the WLA, as expected considering the temporal scope, but even the variants in the WLA are sufficient to make (untrained) modern readers shy away. In both cases, however, the modern equivalent can be found by the use of a spellchecker.

² The spellchecker will be applied to editions as prepared by philologists who will resolve all problems relating to paleography. An unresolved problem is the different treatment of word boundaries, as seen in Figure 3. The spellchecker can be expected to cope with that to a degree, as word boundaries are also a problem in modern spelling.

5. Automatic spelling normalization

Experiments with automatic spelling normalization with a spellchecker have been ongoing at the Árni Magnússon Institute for Icelandic Studies for some years, mainly working on 19th century texts (Svavarsdóttir et al. 2014, Daðason et al. 2014). The spellchecker Skrambi (Daðason 2012) is based on a noisy channel model. Given an unknown word, it will find all similar word forms in the Database of Modern Icelandic Inflection (Bjarnadóttir 2012). These potential corrections are then ranked according to their frequency (as determined from a large corpus) as well as the probability of the specific character edits (i.e. substitutions, deletions, insertions and transpositions) that were required to transform (misspell) the suggested word into the unknown word. Skrambi is used in tandem with other NLP tools for the analysis of Icelandic texts, as in PoS tagging, lemmatization, and compound splitting, but all currently available tools are made to be used on Modern Icelandic only. As NLP tools are generally expensive to make and often rely on extensive data, the advantage of normalizing older texts in order to be able to analyze them and make them searchable by using existing tools is obvious. The crux of the matter is whether the changes in spelling are regular enough to make this possible. This proved to be the case with the 19th century texts referred to above, even though the spelling of some of them was highly eccentric and idiosyncratic. The next step was to experiment with even older texts.

When Skrambi is extended to normalize older texts, it simply treats the older variants as errors and tries to correct them by comparison with modern variants. Skrambi utilizes machine learning techniques, i.e. it learns from the errors it corrects, and access to quantities of variant forms that are attributed to source and thus dated is of immense value. This is where data from both the ONP and the WLA is of great importance, not only in this project but in future work with Icelandic texts from all ages.

6. The dating of words in the WLA

As the WLA is excerpted in the traditional manner of reading and producing dictionary slips of noteworthy examples of usage, the dating of words can

sometimes be quite faulty. To name an example, the headword *eiginmaður* ‘husband’ only has 4 citations in the WLA, all with older (non-standardized) spelling, which in itself makes them noteworthy. The modern spelling does not appear at all under the headword, giving insufficient information of a common noun. The citations in the WLA have recently been made accessible for text search, and there are in fact there 124 additional examples of the modern spelling of *eiginmaður* in the WLA, contained in citations under different headwords. The dating of headwords and variants in the WLA should therefore be taken with a grain of salt, as the traditional method of excerption is not a guarantee of correctly dated words.

The general public in Iceland is not sufficiently well aware of the ONP and of the necessity of consulting both the WLA and the ONP when trying to date words. In Iceland, the WLA sometimes has the immense status of being the definite source of information on a word. People will therefore jump to conclusions, and inane as it may sound, the answer to the question “How old is the word *hundur* ‘dog’?” may very well be “1540”, as the oldest citation of the word in the WLA is indeed from the 1540 New Testament.

Other lacunae in the WLA are more difficult to spot, as seen in Table 3, which shows the examples of headwords from the letter D in the ONP in which the dating in the WLA appears misleading.

<i>Dating</i>	<i>No. of entries</i>	<i>Example</i>	
No date	7	<i>dammur</i> n.masc.	‘dam’
15 th C	2	<i>drykkjarhorn</i> n.neut.	‘drinking horn’
16 th C	280	<i>dauðasvefn</i> n.masc.	‘sleep of death’
17 th C	195	<i>danskur</i> adj.	‘Danish’
18 th C	89	<i>dauðdagi</i> n.masc.	‘day/manner of death’
19 th C	202	<i>Danaveldi</i> n.neut.	‘Kingdom of Denmark’
20 th C	78	<i>dagdrykkja</i> n.fem.	‘day drinking’

Table 3: The dating of words from the ONP in the WLA. Examples from D.

As the ONP shows clearly, the adjective *danskur* ‘Danish’ is not a 17th century word, and *dauðdagi* ‘day of death’ is not from the 18th century. The word *dagdrykkja* (cf. Figures 1 and 2) is an interesting case where the 20th

century word in the WLA may indeed be a ‘new’ word, as the meaning ‘day drinking’ (as a sign of alcoholism) is different from the meaning in the ONP, where it signifies ‘breakfast’ (Orkneyinga saga). The compound is perfectly regularly made, however, then and now, made with the same constituents.

The users of the WLA would be well served if they could access the ONP directly, without technical hitches. To make that possible, the lists of head-words must be connected, and the easiest way is to normalize both to the same standard, which in the case of the WLA is Modern Icelandic.

7. Conclusion

The aim of the approach described in this paper is to maximize the usefulness of linguistic resources and the NLP tools created in the last few years for the analysis of Modern Icelandic texts. Language resources and tools are expensive to make, and in the case of both the ONP and the WLA, the archives have been decades in the making. Both of them are magnificent resources that should be easily accessible to as many people as possible, in as many ways as possible. If existing NLP tools can be adapted to cope with older forms of Icelandic, the general public would gain better access to the history of individual words, and scholars (i.e., philologists, lexicographers, linguists, historians, etc.) would be able to utilize the texts to a fuller extent. By the anchoring of spelling variants and lemmas, a new resource for research would be created. The benefits of the normalization would be better access to Icelandic vocabulary throughout history, in lexicography, in research on the language in general, and for the general public, as in the creation of search engines coping with Icelandic of all ages.

Bibliography

Dictionaries

Cleasby, Richard & Guðbrandur Magnússon (1874): *An Icelandic-English Dictionary*. Oxford: Clarendon Press.

Fritzner, Johan (1867): *Ordbog over Det gamle norske Sprog*. Kristiania: Feilberg & Landmark.

ONP = *Dictionary of Old Norse Prose* [*Ordbog over det norrøne prosasprog*]: <onp.hum.ku.dk>.
WLA = The Written Language Archive: <arnastofnun.is/page/gagnasofn_ritmalssafn>.

Other references

- Bjarnadóttir, Kristín (2012): The Database of Modern Icelandic Inflection. In: *LREC 2012 Proceedings: Proceedings of “Language Technology for Normalization of Less-Resourced Languages”*, *SaLTMiL 8 – AfLaT 2012*, 13-18.
- Daðason, Jón Friðrik (2012): *Post-Correction of Icelandic OCR Text*. <hdl.handle.net/1946/12085>.
- Daðason, Jón Friðrik, Kristín Bjarnadóttir & Kristján Rúnarsson (2014): The Journal *Fjölur* for Everyone: The Post-Processing of Historical OCR Texts. In: *Proceedings of Language Resources and Technologies for Processing and Linking Historical Documents and Archives – Deploying Linked Open Data in Cultural Heritage*, 55-62. LRT7HDA. LREC 2014, Reykjavík.
- Gottskálksson, Oddur (1540): *Nýja testamentið*. Roskilde.
- Gunnlaugsson, Guðvarður Már (ed.) (2000): *Reykjaholtsmáldagi*. Reykholt, Reykholtskirkja, Snorrastofa.
- Jóhannsson, Ellert Þór & Simonetta Battista (2016): *Ordbog over det norrøne prosasprog online – struktur og brug*. This publication, xx-xx.
- Svavarsdóttir, Ásta, Sigrún Helgadóttir & Guðrún Kvaran (2014): Language Resources for Early Modern Icelandic. In: *Proceedings of Language Resources and Technologies for Processing and Linking Historical Documents and Archives – Deploying Linked Open Data in Cultural Heritage*, 19-25. LRT7HDA. LREC 2014, Reykjavík.

Kristín Bjarnadóttir
research lecturer
The Árni Magnússon Institute for Icelandic Studies/Háskóla Íslands
Laugavegi 13
IS-101 Reykjavík
kristinb@hi.is

LEXIN i Norge

– Hva sier brukerne?

Tove Bjørneset

The original Swedish dictionary series *LEXIN* was launched in the 1980's, designed specifically for minority language immigrants living in Sweden. In 1996, The Norwegian Ministry for Education, Research and Church Affairs initiated a corresponding Norwegian project at The Norwegian Term Bank. The project has since continued in the same environment. At present it is a part of the university-owned company Uni Research, and it is owned and funded by The Norwegian Directorate for Education and Training (Utdanningsdirektoratet). The Norwegian series consists of dictionaries from Norwegian Bokmål and Nynorsk into 16 target languages, and more languages are forthcoming. Since the launch in 2002 the dictionary users have been able to contact the project group directly through the project's website. The user dialogue has proven to be of great value. Two user surveys have also contributed to improve our knowledge about the user needs and about how they benefit from using Norwegian LEXIN.

1. Utgangspunkt og bakgrunn

Innvandringen på 1970- og 1980-tallet gjorde Sverige til et flerkulturelt land. Det skapte et behov for nytenkning rundt språkundervisning og ordbøker til minoritetsspråk. I 1977 tok den svenske Skolöverstyrelsen initiativ til ordbokprosjektet LEXIN (*Lexikon för invandrare*). Göteborgs universitet fikk i oppdrag å utarbeide det svenske ordbokunderlaget (Gellerstam 1999:3). De illustrerte LEXIN-ordbøkene var innholdsrike og enkle å bruke, og ble godt likt av både lærere og elever på ulike språknivåer.

Danmark, Island og Norge takket senere ja til Sveriges tilbud om gratis lån av det maskinlesbare LEXIN-materialet med oversettelser til flere språk.

I Norge var det *Det norske kirke-, undervisnings- og forskningsdepartementet* (KUF) som i 1996 rettet en forespørsel til tidligere *Norsk termbank* i Bergen om et pilotprosjekt i det flerfaglige miljøet der. Prosjektet er fortsatt i det samme fagmiljøet, i dag en del av det universitetside forskningsseksjonen *Uni Research*.

Det er utgitt flere artikler nasjonalt og internasjonalt om arbeidet i norsk LEXIN (blant andre Bjørneset, T. 2005; 2002; 2001; 1999) og om prosjektet sett utenfra (blant andre Pálfi, L.L. & S. Tarp 2009 og Hovdenak, M. 2008). I 2008 fikk LEXIN-prosjektet Rosingprisen i språk fra Den norske dataforening i samarbeid med Språkrådet.

2. Norsk LEXIN i dag

Norsk LEXIN er i likhet med det svenske moderprosjektet et ikke-kommerisielt prosjekt. Det er eid og finansiert av den norske stat ved Utdanningsdirektoratet (Udir) siden 1998. Udir definerer minoritetsspråklige elever i grunnopplæringen som primærmålgruppen. De kartlegger målgruppens sammensetning og behov løpende, og avgjør blant annet på dette grunnlaget hvilke språk og teknologisk videreutvikling som skal prioriteres hvert prosjektår. Prosjektgruppen har på sin side ansvaret for praktisk planlegging og gjennomføring av alt språkfaglig og teknisk utviklings- og vedlikeholdsarbeid, inklusive opplæring og oppfølging av prosjektmedarbeidere.

2.1. Oversikt over ordbøker og andre språkressurser

Ved utgangen av første halvår 2015 var det utgitt 35 nettbaserte og 6 trykte ordbøker på totalt 18 språk i den norske serien, mens ytterligere to nettbaserte og to trykte ordbøker er i arbeid.¹

¹ De nettbaserte ordbøkene ligger fritt tilgjengelig her: <lexin.udir.no/>.

2.1.1. Nettbaserte ordbøker

Alle de nettbaserte ordbøkene har både bokmål og nynorsk som søkespråk.

Bokmålsordbøker	Nynorskordbøker
<i>Bokmålsordbok</i> (36 000, 2002)	<i>Nynorskordbok</i> (36 000 ord, 2010)
<i>Bokmål–nynorsk</i> (36 000, 2010)	<i>Nynorsk–bokmål</i> (i arbeid)
<i>Bokmål–arabisk</i> (11 000, 2007)	<i>Nynorsk–arabisk</i> (11 000 ord, 2010)
<i>Bokmål–dari</i> (11 000, 2012)	<i>Nynorsk–dari</i> (11 000 ord, 2012)
<i>(Bokmål–engelsk</i> (36 000, 2002)) ²	<i>Nynorsk–engelsk</i> (36 000 ord, 2010)
<i>Bokmål–kurdisk kurmanji</i> (11 000, 2004)	<i>Nynorsk–kurdisk kurmanji</i> (11 000, 2010)
<i>Bokmål–kurdisk sorani</i> (11 000, 2004)	<i>Nynorsk–kurdisk sorani</i> (11 000, 2010)
<i>Bokmål–persisk</i> (11 000, 2007)	<i>Nynorsk–persisk</i> (11 000, 2010)
<i>Bokmål–polsk</i> (11 000, 2010)	<i>Nynorsk–polsk</i> (11 000, 2010)
<i>Bokmål–russisk</i> (11 000, 2010)	<i>Nynorsk–russisk</i> (11 000, 2010)
<i>Bokmål–somalisk</i> (11 000, 2005)	<i>Nynorsk–somalisk</i> (11 000, 2010)
<i>Bokmål–tagalog</i> (11 000, 2012)	<i>Nynorsk–tagalog</i> (11 000, 2010)
<i>Bokmål–tamil</i> (7 500, 2002)	<i>Nynorsk–tamil</i> (7 500, 2010)
<i>Bokmål–thai</i> (11 000, 2010)	<i>Nynorsk–thai</i> (11 000, 2010)
<i>Bokmål–tigrinja</i> (11 000, 2007)	<i>Nynorsk–tigrinja</i> (11 000, 2010)
<i>Bokmål–tyrkisk</i> (11 000, 2007)	<i>Nynorsk–tyrkisk</i> (11 000, 2010)
<i>Bokmål–urdu</i> (11 000, 2009)	<i>Nynorsk–urdu</i> (11 000, 2009)
<i>Bokmål–vietnamesisk</i> (11 000, 2012)	<i>Nynorsk–vietnamesisk</i> (11 000, 2012)
<i>Bokmål–litauisk</i> (11 000, i arbeid)	<i>Nynorsk–litauisk</i> (11 000, i arbeid)
<i>Bokmål–burmesisk</i> (11 000, i arbeid)	<i>Nynorsk–burmesisk</i> (11 000, i arbeid)

2.1.2. Trykte ordbøker og språkressurser

<i>Norsk–tamilsk illustrert ordbok</i> (7 500, 2004)
<i>Illustrasjonshefte bokmål og nynorsk</i> (1 920, 2004)
<i>Norsk–kurdisk sorani illustrert ordbok</i> (11 000, 2005)
<i>Norsk–kurdisk kurmanji illustrert ordbok</i> (11 000, 2007)
<i>Norsk–somalisk illustrert ordbok</i> (11 000, 2007, utsolgt, andreutgaven i arbeid)
<i>Norsk–tigrinsk</i> ³ <i>illustrert ordbok</i> (11 000, 2011)
<i>Norsk–thai illustrert ordbok</i> (11 000, 2013)
<i>Norsk–dari illustrert ordbok</i> (11 000, i arbeid)

² Ble besluttet fjernet av Udir høsten 2009, etter krav fra kommersielle forlag.

³ Det norske språknavnet er i dag *tigrinja*, men dette var ikke normert i 2011.

2.1.3. Bildetema

Bildetema er en interaktiv bildeordbok med rundt 1 900 ord fordelt på 31 temaer og med naturlig tale på per i dag 36 språk. Den norske versjonen bygger på *Bildteman*, som ble laget i Sverige tidlig på 2000-tallet. Ord- og bildematerialet er hentet fra LEXIN-ordbøkene, men er presentert på en måte som gjør det særlig godt egnet for små barn. Utarbeidelsen av nye språkversjoner og teknisk videreutvikling skjer i nært samarbeid med prosjektgrupper i Sverige og Danmark.

2.2. LEXINs brukergrensesnitt

Helt fra prosjektstarten i 1997 har prosjektgruppen hatt ansvaret for skreddersynging og vedlikehold av redigeringsgrensesnittet for redaktører, og siden 2002 også for sluttbrukerne. Fra 1998 har all redigering foregått via nettet.

I 2010 ble det lagt til rette for fritekstsøk på bokmål og nynorsk. Det at brukerne nå kan søke på blant annet bøyde former av oppslagsordet og i en del tilfeller få flere hundre treff på søkeordet i ulike kontekster, gir en langt mer effektiv utnyttelse av materialet og muligheter for større bruksvariasjon.

Etter ønske fra Udir implementerte vi i 2011 talesyntese for alt norsk-språklig innhold.⁴ Udir valgte en synteseløsning for å kunne gi utviklingen av ordbøker til flere målpråk høyest mulig prioritet. Ifølge referansegruppen for LEXIN kan det for mange minoritetsspråklige elever i grunnopplæringen være enklere å oppfatte de ulike lydene og lydkombinasjonene i talesyntese enn i naturlig tale. En annen og åpenbar fordel ved talesyntese er at vi kan gjøre endringer i databasen uten å måtte administrere nye lydinnspillinger.

2.3. Brukerstatistikk

Systematisk høsting, tolking og formidling av konsistent brukerstatistikk kan være en krevende øvelse, og opp gjennom årene har vi både testet og brukt ulike statistikkprogrammer. Fra og med første halvår 2015 har vi valgt å

⁴ Synteseløsningen LeseWEB eies og driftes av firmaet VoiceASP, mens de fire stemmene Udir har valgt å bruke for henholdsvis bokmål og nynorsk er laget av Acapela Group og Lingit.

bruke *Piwik*, som oppgir følgende besøkstall basert på forespørsler fra unike IP-adresser i perioden:

- Gjennomsnittlig antall besøk per dag: 6 662
- Antall besøk dagen med størst trafikk: 11 986

Tabellen under viser antall ganger brukere har skrevet inn et søkeord og trykket på søk-knappen i en av de nettbaserte ordbøkene:

- Januar: 1 881 440
- Februar: 2 314 728
- Mars: 1 647 620
- April: 1 389 357
- Mai: 1 423 696
- Juni: 973 648

Totalt: 9 630 489

3. Dialogen med brukerne

De norske LEXIN-ordbøkene har vært gratis og fritt tilgjengelig på nettet siden de tre første⁵ ble publisert høsten 2002. Like lenge har brukerne kunnet kommunisere direkte med oss via en e-postlenke på prosjektets hjemmeside. Vi la til rette for dette for at det skulle være enklest mulig for brukerne å melde fra om tekniske problemer eller feil. Det viste seg imidlertid raskt at kanalen skulle bli like mye brukt til andre typer innspill, men også til dialog.

Jevnlig kontakt og dialog med brukerne har gitt oss verdifull kunnskap om både dem og om noen av de utfordringene de møter i Norge, kunnskap vi nok vanskelig kunne fått på annen måte. Brukergruppen er heterogen og internasjonal, men de fleste som tar kontakt er elever, lærere eller studenter bosatt i Norge. Mange trenger hjelp til å løse konkrete tekniske problemer, som nedlasting og installasjon av fonter, men vi får også mange forespørsler om nye målspråk og ordbøker, og forslag til ny funksjonalitet.

⁵ *Bokmålsordbok, Bokmål–engelsk ordbok og Bokmål–tamil ordbok.*

Noen forespørsler gir oss en mulighet til å stille spørsmål og få svar fra språkkyndige personer med svært varierende språkbakgrunn. I et par tilfeller har slik dialog fått direkte innvirkning på Udirs prioriteringer. Blant annet var det gjentatte krav om nynorsk fra minoritetsspråklige elever bosatt i nynorskkommuner som førte til at nynorsk i 2010 ble implementert som kilde-språk i samme omfang som bokmål. Tilsvarende var det mange forespørsler fra dari-talende som til slutt førte til at det ble laget en nettbasert ordbok fra norsk til dari. Denne skal også utgis i trykt utgave.

En del brukere ber om hjelp til å skaffe seg oversikt over og eventuelt også bestille ordbøker på det nasjonale eller internasjonale markedet. Det kan skyldes at det aktuelle språket mangler i LEXIN eller at de trenger større eller mer avanserte ordbøker. Noen ønsker hjelp til å lære norsk. Det har vi ikke kapasitet til, men vi prøver å hjelpe dem videre med blant annet informasjon om ulike undervisningsaktører. Vi blir også fra tid til annen kontaktet av personer som ønsker å hjelpe andre med skolearbeidet eller med å lære norsk, men som trenger starthjelp for å bli i stand til dette.

Siden 2002 har vi flere ganger årlig møtt på og snakket med brukere på konferanser og seminarer om språk, språkundervisning, mangfold, inkludering og bruk av IKT i utdanning og læring. I samme periode har vi holdt mange innlegg, presentasjoner og bokutstillinger på de samme arenaene og møteplassene. På bokutstillingene møter vi brukere med til dels svært ulikt utgangspunkt for sin bruk av LEXIN. Lærere for introduksjonsklasser og for enslige mindreårige helt uten norskkunnskaper har fortalt oss at LEXIN og Bildetema er populære og gir godt læringsutbytte. Vi har også fått vite at utenlandske innsatte i norske fengsler bruker LEXIN for å lære seg norsk.

4. Brukerundersøkelser

Det er gjennomført to brukerundersøkelser av norsk LEXIN. Den første var en kvalitativ undersøkelse av mindre omfang som ble gjennomført i 2006 i samarbeid mellom Udir og prosjektgruppen. Den andre ble gjennomført i 2011 av Rambøll Management Consulting på oppdrag fra Udir.⁶

⁶ Rambøll-undersøkelsen ble publisert i januar 2012 og er tilgjengelig her: www.udir.no/globalassets/upload/5/lexin.pdf.

I den første brukerundersøkelsen svarte de fleste respondentene at de mener LEXIN egner seg for bruk på flere måter, øker læringsmotivasjonen og oppleves som nyttig av brukere på ulike språknivåer. Mange mente at brukervennligheten gjør LEXIN nyttig for brukere som ikke ville hatt særlig utbytte av konvensjonelle ordbøker. De fleste var fornøyd med grensesnittet.

4.1. Rambøll-undersøkelsen

Udirds målsetting med Rambøll-undersøkelsen i 2011 var å innhente informasjon om hvordan LEXIN blir brukt og om brukernes erfaringer, samt å få innspill til videreutvikling fra blant andre brukerne av LEXINs nettsted, skoleledere, lærere, forlagsfolk, ansatte i Språkrådet og medlemmene i prosjektets egen referansegruppe. Det skulle også ses nærmere på ulike brukergruppers bruksomfang og brukernes tilfredshet med brukergrensesnittet. Udir ville ha en kombinasjon av kvantitative og kvalitative data for å nå et så bredt utvalg som mulig. Datainnsamlingen gikk både i bredde og dybde gjennom tre brukerundersøkelser og to fokusgruppeintervjuer med oppfølgende kvalitative intervjuer. Undersøkelsens kvantitative del bestod av en brukerundersøkelse via et nettbasert spørreverktøy på LEXINs nettside og en brukerundersøkelse blant lærere i NAFOs⁷ fokusvirksomheter og et utvalg skoleledere. Den nettbaserte undersøkelsen var rettet mot alle brukere av LEXINs nettsted i perioden 19.10.–11.11.11 og forelå på seks språk. Av totalt 8 096 unike IP-adresser ble det avgitt fullstendige svar fra 1 116 (norsk (622), engelsk (275), polsk (93), arabisk (70), somali (54) og urdu (2)).

Undersøkelsen viser at brukergruppen er så godt som kjønnsbalansert, med to prosent overvekt av mannlige brukere, men ellers heterogen i svært mange henseender. Den yngste brukeren var 6 år, den eldste 81, og gjennomsnittsalderen var 31 år. Hele 71 prosent svarte at de ikke er født i Norge, mens 9 prosent svarte at de var bosatt utenfor Norge. De som bruker LEXIN oftest, har bodd i Norge i 2–5 år. LEXIN brukes mye, særlig av elever med særskilte behov. Rundt 80 prosent bruker LEXIN hjemme, og rundt 40 prosent på skolen. Rundt 35 prosent bruker LEXIN daglig, mens rundt 50 prosent bruker LEXIN to ganger i uken eller mer. Mange foreldre svarer at

⁷ Nasjonalt senter for flerkulturell opplæring ved Høgskolen i Oslo og Akershus: <nafo.hioa.no/>

de bruker LEXIN for å hjelpe barna med skolearbeidet. Av lærerne oppgir 80 prosent at de bruker LEXIN, særlig i grunnleggende norskundervisning for minoriteter. De fleste lærerne som bruker LEXIN i undervisningen, har mer enn 10 års arbeidserfaring. Over 50 prosent av disse jobber i barneskolen, mens 12 prosent jobber i voksenopplæringen.

Rambøll konkluderer med ”stor grad av tilfredshet” når det gjelder LEXIN som læringsverktøy og ”behovet LEXIN dekker” hos brukergruppene. Over 40 prosent av de spurte mener at LEXIN dekker behovet for aldersgruppen 6–16 år. Det er også stor grad av enighet om at både de nettbaserte og trykte ordbøkene er tilpasset de ulike brukergruppenes behov og språknivå. Dette understøttes av svarene fra lærere og skoleledere, som mener at LEXIN er et enkelt verktøy å bruke og at de vil anbefale det som et språklig hjelpemiddel i opplæringen. Det er stor grad av enighet blant alle brukergrupper (81 prosent er *helt enig* eller *enig*) om at LEXIN bidrar til å gjøre det enklere å lære norsk og til å øke læringsutbyttet.

Rambøll konkluderer med ”stor grad av tilfredshet” også når det gjelder teknisk løsning og brukergrensesnitt. Elever og lærere mener at både de nettbaserte og trykte ordbøkene er ”enkle å bruke og slå opp ord i”. Flertallet mener også at det er enkelt å navigere rundt på nettstedet.

Ifølge den nettbaserte undersøkelsen mener 36 prosent av brukerne at det er behov for flere språk i LEXIN. Rundt 25 prosent av disse ønsker å få bokmål–engelsk-ordboken tilbake, eventuelt en énspråklig engelskordbok som erstatning. ”Svært mange” ønsker språk Udir hittil ikke har vurdert som aktuelle, som fransk, tysk og spansk, flere østeuropeiske språk, hindi og kinesisk. Videre er det mange som ønsker flere illustrasjoner og illustrasjonstemaer og det å kunne søke begge veier i de nettbaserte ordbøkene. En del etterlyser også fagord og faguttrykk, samt flere trykte ordbøker.

Rambøll-undersøkelsens kvalitative del bestod av såkalte *fokusgruppeintervjuer* med blant andre ansatte i Språkrådet, NAFO, VOX og ulike forlag, og med LEXINs referansegruppe. Ifølge flere av disse respondentene er det behov for flere språk i LEXIN, og da først og fremst flere østeuropeiske språk i tillegg til pashto, vietnamesisk og kinesisk. Mange mener også at det er behov for å utvide vokabularet og å kunne søke begge veier i de tospråklige ordbøkene. Innlemmelse av fagord og faguttrykk er også noe flere mener vil være nyttig for elever i yrkesfaglige utdanningsprogrammer i videregående opplæring. De begrunner det med at eksisterende språklige

hjelpemidler er for dårlig tilpasset mange elevers språknivå. En del lærere ønsker flere trykte ordbøker, fordi de mener disse er enklere å bruke i undervisningen, og da særlig i individuell undervisning.

Rambøll konkluderer med ”en stor grad av tilfredshet” når det gjelder LEXIN som læringsverktøy og ”behovet LEXIN dekker”. Både elever og lærere mener at de nettbaserte og trykte ordbøkene er ”enkle å bruke og slå opp ord i”. Mange lærere svarer at de oppnår gode resultater når de bruker LEXIN hos elever med særskilte behov. Det er også stor grad av enighet om at de nettbaserte og trykte ordbøkene er tilpasset ulike brukergruppers behov og språknivå. Dette bekreftes av svarene fra lærere og skoleledere, som mener at LEXIN er et språkverktøy som er enkelt å bruke og som de vil anbefale overfor andre lærere. Funnene i den kvalitative undersøkelsen understøttet funnene i den kvantitative:

- LEXIN er unik i sin form og dekker et viktig behov som ikke ivaretas andre steder.
- LEXIN er et ”svært nyttig verktøy for elever og andre som vil lære seg norsk”.
- LEXIN er nyttig for lærere som underviser i morsmål eller i grunnleggende norsk for minoritetsspråklige.

5. Sluttkommentar

Kontakten med brukerne har vist seg å bli mer verdifull og nyttig for oss enn vi kunne forutse da vi la til rette for direkte dialog gjennom prosjektets hjemmeside ved lanseringen av de første ordbøkene i 2002. Vi har fått innsikt og kunnskap på ulike områder som vi vanskelig kunne fått på annen måte. Brukerkontakten og de to brukerundersøkelsene har også gitt oss et mer nyansert og sammensatt bilde av brukerne og noen av deres utfordringer og behov i det norske samfunnet. For oss er det særlig gledelig og motiverende at Rambøll-undersøkelsen avdekker så stor enighet blant alle brukergrupper om at LEXIN både bidrar til å gjøre det enklere å lære norsk og til å øke læringsutbyttet hos elever med særskilte behov.

Innspill fra brukerne har allerede fått sette sitt preg på LEXIN og Bilde-tema, og de vil være viktige også i fortsettelsen av det norske prosjektet.

Etter den siste brukerundersøkelsen har vi gitt ut en nettbasert norsk–vietnamesisk ordbok, og nå lager vi tilsvarende ordbøker til burmesisk og litauisk. Vi arbeider også med å gjøre alle tospråklige nettordbøker søkbare begge veier.

Ifølge de to brukerundersøkelsene og andre innspill fra brukerne tåler den snart 40 år gamle og banebrytende svenske LEXIN-modellen tidens tann. Den er fortsatt en populær og effektiv språkressurs for en stadig voksende og svært heterogen gruppe innvandrere, asylsøkere og flyktninger i Norge.

Litteratur

Ordbøker

De nettbaserte ordbøkene i norsk LEXIN: <lexin.udir.no/>

Annen litteratur

Bjørneset, Tove (2005): The NORDLEXIN-N Dictionary Project: Part II. I: Henrik Gottlieb, Jens Erik Mogensen & Arne Zettersten (red.): *Symposium on Lexicography XI. Proceedings of the Eleventh International Symposium on Lexicography May 2–4, 2002 at the University of Copenhagen* (Lexicographica Series Mayor 115), 133-145. Tübingen.

Bjørneset, Tove (2002): The Dictionary Project NORDLEXIN-N. I: Henrik Gottlieb, Jens Erik Mogensen & Arne Zettersten (red.): *Symposium on Lexicography X. Proceedings of the Tenth International Symposium on Lexicography May 4–6, 2000 at the University of Copenhagen* (Lexicographica Series Mayor 109), 31-43. Tübingen.

Bjørneset, Tove (2001): Introduksjon til ordboksprosjektet NORDLEXIN-N. I: Martin Gellerstam, Kristinn Jóhannesson, Bo Ralph & Lena Rogström (red.): *Rapport från Konferens om lexikografi i Norden, Göteborg 26–29 maj 1999. Nordiska studier i lexikografi 5* (Meijerbergs arkiv för svensk ordforskning 2001 (27)), 44-53. Göteborg.

Bjørneset, Tove (1999): Ordboksprosjektet NORDLEXIN-N. I: *Lexico-Nordica* 6, 35-45.

Nordiske Studier i Leksikografi 13 (2016): 89-99

- Bjørneset, Tove & Ásta Svavarsdóttir (2005): Nord-Lexin. I: Vatvedt Fjeld, Ruth & Dagfinn Worren (red.): *Nordiske studiar i leksikografi 7. Rapport frå Konferanse om leksikografi i Norden, Volda 22.–24. mai 2003*, 52-66. Oslo.
- Gellerstam, Martin (1999): LEXIN – lexikon för invandrare. I: *LexicoNordica* 6, 3-17.
- Hovdenak, Marit (2008): Dei norske LEXIN-ordbøkene. I: *LexicoNordica* 15, 219-234.
- Hult, Ann-Kristin, Sven-Göran Malmgren & Emma Sköldberg (2010): Lexin – A report from a recycling lexicographic project in the North. I: *Proceedings of the XIV Euralex International Congress, Leeuwarden, 6–10 July 2010*, 800-809.
- Hult, Ann-Kristin (2008): Från ord till handling. En studie i ordboks-användning på nätet. I: *Nordiska studier i leksikografi. Rapport från 9. konferens om leksikografi i Norden, Akureyri 22.–26. maj 2007*, 243-257. Reykjavik.
- Hult, Ann-Kristin (2008): Användarna bakom loggfilerna. Redovisning av en webbenkät i Lexin online Svenska ord. I: *LexicoNordica* 15, 73-91.
- Malmgren, Sven-Göran (1999): De svenska LEXIN-ordböckerna ur ett (ställföreträdande) användarperspektiv. I: *LexicoNordica* 6, 79-90.
- Pálfi, Loránd-Levente & Sven Tarp (2009): Lernerlexikographie in Skandinavien – Entwicklung, Kritik und Vorschläge. I: *Lexicographica. International Annual for Lexicography* 25/2009, 135-154. Tübingen.
- Rambøll Management Consulting (på oppdrag for Utdanningsdirektoratet) (2011): *Brukerundersøkelse av LEXIN – ordbøker for minoritetsspråk-lige*: <www.udir.no/globalassets/upload/5/lexin.pdf>

Tove Bjørneset
 prosjektleder, cand.philol.
 Uni Research, Gruppe for språk og språkteknologi
 Postboks 7810
 NO-5020 Bergen
 tove.bjorneset@uni.no

Indexeringen av SAOB

Erik Bäckerud

We are creating a new digital version of the Swedish Academy dictionary, SAOB. We have previously scanned the printed originals to obtain a correct text in digital format and we are presently attempting to automatically identify important structures in the articles. The structures we identify include title, part of speech, etymology and the division of definitions among others. This paper will give details of our work and a review of how well we are succeeding and what difficulties we have encountered.

1. En digital SAOB

1.1. Bakgrund

Svenska Akademiens ordbok (SAOB) har digitaliserats en gång tidigare, den gången av en grupp knuten till dåvarande avdelningen Språkdata vid Institutionen för svenska språket, Göteborgs universitet (OSA 1996). Vi har tack vare nydigitaliseringen fått ned antalet fel i texten väsentligt, både vad gäller tecken och stilar.

De primära målen med indexeringsarbetet är:

- Att göra en webbpresentation av SAOB som uppfyller höga krav på läsbarhet och korrekthet och som riktar sig såväl till fackmän som till allmänheten.
- Att skapa indata till ett framtida redigeringsystem då ordboken skall uppdateras.

1.2. Grunderna

Skapandet av en ny digital version av SAOB företas i två steg. Det första steget, att skanna all text till en teckenrätt digital version, har jag beskrivit i

ett tidigare arbete (Bäckerud 2014). Detta arbete är nu avslutat och utfallet blev mycket gott.

I det andra steget, som beskrivs i denna artikel, utvecklar vi automatiska metoder för att identifiera information om strukturer som finns implicit i artikeltexterna och göra dessa explicita. Den första fasen av indexeringsarbetet är klar t.o.m. band 36. Arbetet fortsätter nu med kvalitetsförbättringar och indexering av nyare material.

De strukturer vi indexerat är (jfr Lundbladh 1992):

- 1) Uppslagsord med biformer
- 2) Sammansättningar
- 3) Avledningar
- 4) Särskilda förbindelser (dvs. förbindelser av verb och betonad partikel)
- 5) Ordklass
- 6) Etymologiparentes (i huvudet)
- 7) Formparentes i huvudet (dvs. stavningsformer m.m.)
- 8) Moment I, II osv., A, B osv. och 1, 2 osv.
- 9) Sekundära sammansättningar och avledningar
- 10) Översikt
- 11) Hänvisningsartiklar

Utöver ovan uppräknade punkter har vi också till punkt 8 indexerat alla undermoment ned till den lägsta nivån; α' , β' osv. och till punkt 9 har vi även identifierat sekundära särskilda förbindelser vilka förekommer enstaka gånger i ordbokstexten.

1.2.1. Indata

Det material vi haft tillgång till under projektet är texten till SAOB band 1 t.o.m. 36 omfattande uppslagsorden A till UTSTÄDES. De olika banden finns tillgängliga i olika format eftersom redaktionen använt olika tekniska lösningar vid olika tider. Band 1–31 finns i tryckt form på papper och har skannats, medan de senare banden finns i olika ordbehandlarformat.

Totalt har vi därmed 36 band med sammanlagt 195 miljoner skrivtecken fördelade på 84.000 artiklar, varav 54.000 är definitionsartiklar. Totalt har vi identifierat cirka 567.000 uppslagsformer inklusive hänvisningar och biformer.

2. Indexeringsarbetet

2.1. Om artiklarnas uppbyggnad

Här nedan följer exempel på hur artiklar i SAOB är uppbyggda och hur de olika strukturer vi indexerar ser ut.

2.1.1. Huvudet

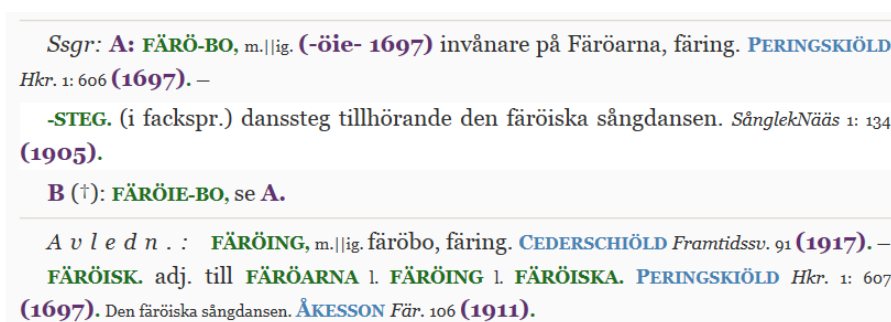
Flera av punkterna som listas i 1.2 ovan kan man få fram genom att analysera den inledande delen av artikeln, det så kallade huvudet. I nedanstående figur har jag markerat uppslagsordet med biformer (dvs. FLÅTE och FLOTTA), ordklass med homografsiffra, formparentes och etymologi-parentes.

FLOTTE *flot*³ *e*², sbst.², l. (numera bl. starkt bygdemålsfärgat) **FLÅTE** *flå*³ *te*², r. l. m.; best. **-en** ((†) **-flått**n **RÅLAMB** 10: 42 (1691)); pl. **-ar**; äv. (numera bl. bygdemålsfärgat samt i Finl.) **FLOTTA** *flot*³ *a*², sbst.², r. l. f.; best. **-an**; pl. **-or** ((†) **-er** (i vissa fall möjl. att hänföra till sg. **flotte**) **2Krön.** 2: 16 **Bib.** (1541), **HALLENBERG Hist.** 5: 260 (1796)). (**flott-** (**flått-**) c. 1540 osv. **flot-** 1538–1734. **flåt(h)-** 1588–1889. **-e** 1538 osv. **-a** c. 1540 (*oblik kasus*)–1695 (*nom.*) –1916 (: *flottor*)) [*fsv. floti*, motsv. d. *flaade*, isl. *floti*, feng. *flota*, eng. *float*, bildat på det svaga rotstadiet till **FLYTA**; jfr **FLOTA**, sbst.¹, **FLOTT**, sbst.^{1–2}, **FLOTT**, adj.^{1–2}, **FLOTTA**, sbst.¹, **FLOTTA**, v.^{1–2}]

Figur 1: Huvudet i artikeln FLOTTE sbst.².

2.1.2. Övriga uppslagsformer

I SAOB förekommer särskilda förbindelser (verb med partikel), sammansättningar och avledningar i så kallade ramsor. I exemplet nedan ser vi dels en sammansättningsramsa med två underordnade ramsor med olika förled, dels en avledningsramsa med ett par avledningar.



Figur 2: Sammansättnings- och avledningsramsor.

I exemplet identifieras sammansättningarna FÄRÖ-BO, FÄRÖ-STEG och FÄRÖIE-BO samt avledningarna FÄRÖING och FÄRÖISK.

2.1.3. Sekundära konstruktioner och biformer

I SAOB kan både sammansättningar och avledningar i sin tur ha sammansättningar och avledningar. Dessa kallas då sekundära sammansättningar respektive avledningar. Här nedan visas ett exempel från artikeln TRO sbst., under sammansättningen TRO-LOVEN där det förekommer två sekundära sammansättningar trolovan-spillare och trolovan-öl.

-LOVEN l. **-LOVE** l. **-LOVAN**. [fsv. *trolovan*; formen **-lovan** snarast vbalsbst. till **-lova**] (†) (högtidligt l. edligt) löfte om trohet; äv. speciellare: trolovning. Ther om och alt annat wij eder formane, på samma hulskap och troloffue som j med oss hans nade haffue rethuiselige tilsagt, atj bliffue ther faste vtinnan och retther eder sielffue j alle motte. *G1R* 4: 434 (1527). Ther trolofwan är skeedd medh lagha gåfwor och närwarande witne och sedan är sängelagh tilkommit, sliik handel moste man räkna gill och rätt in för Gudh. *RUDBECKIUS Kyrkiost.* 22 (c. 1635). *ALMQVIST Amor.* 309 (1822, 1839). Ssgr (†): **trolovan-** l. **troloves-** l. **trolovs-spillare**. person som omintetgjort trolovning. *VDAkt.* 1656, nr 112. **-öl**. trolovningsöl. *BUREUS Suml.* 55 (c. 1600). *HammarkDomb.* 16/9 1623. —

Figur 3: Exempel från artikeln TRO sbst.¹

Vi ser här också flera exempel på sammansättningar med biformer. TRO-LOVEN har två biformer: TRO-LOVE och TRO-LOVAN och den sekundära sammansättningen trolovan-spillare har två biformer: troloves-spillare och trolovs-spillare.

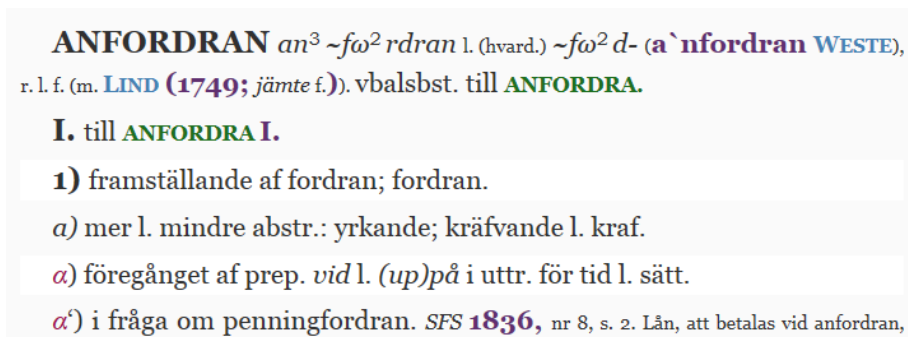
2.1.4. Moment

Betydelsebeskrivningarna i SAOB är indelade i numrerade moment vilka förekommer hierarkiskt i inte mindre än fem olika nivåer.

1) Romerska (över-)moment	I, II, III, ...
2) Arabiska siffror (huvudmoment)	1, 2, 3, ...
3) Undermoment	a, b, c, ...
4) Grekiska undermoment	α , β , γ , ...
5) Grekiska prim-undermoment	α' , β' , γ' , ...

Till ovanstående kommer att momenten på de två högsta nivåerna kan vara avskilda med grupperingsbokstäver som skrivs A, B, C osv. Dessa bryter dock inte numreringen utan finns endast för att förtydliga sambandet mellan olika intilliggande moment.

I nästa exempel visas början av en artikel som innehåller alla fem nivåerna av momentnumrering.



Figur 4: Artikel med numrerade moment i fem nivåer.

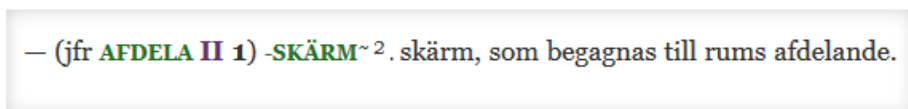
2.2. Svårigheter

Arbetet med indexeringen har överlag gått bra men naturligtvis har vi stött på svårigheter. Ett återkommande problem har varit fel i den ursprungliga texten. Obalanserade parenteser och utelämnade bindestreck får obönhörligt indexeringsprogrammet att gå vilse.

Även när texten är korrekt finns det svåra fall där indexeringen kan gå fel. Några exempel ges nedan.

2.2.1. Exempel

Momentsiffror i SAOB skrivs som en siffra följd av en parentes, **1)**, eller som en bokstav följd av parentes, *a)*. En hänvisning till ett visst moment i en annan artikel som står inom parentes kan lätt förväxlas med en momentsiffra. I nedanstående exempel är **1)** inte början av ett moment.



Figur 5: En hänvisning inom parentes.

För att undvika detta fel måste programmet söka bakåt i texten och se att det finns en oavslutad vänsterparentes 16 tecken tidigare. Ibland måste man se

ännu längre än så för att förvissa sig om att det inte handlar om en hänvisning.

I nästa exempel har redaktören inte skrivit ut de fulla formerna i avledningen FÖRSTÄRKING. Det som avses är ju att de tre orden FÖRSTÄRKING, FÖRSTÄRKNING och FÖRSTÄRKUNG alla hänvisar till artikeln FÖRSTÄRKNING.

FÖRSTÄRKING, -NING, -UNG, se FÖRSTÄRKNING.

Figur 6: Svårtolkad hänvisning.

Svårigheten här består i att programmet inte kan urskilja vilken del av ordet FÖRSTÄRKING som skall föras ihop med respektive efterled. Det blir därmed omöjligt att återskapa de korrekta formerna automatiskt.

2.3. Utfall

Under hösten 2014 genomförde jag och Erik Norberg, en timanställd student, ett större test där vi korrekturläste en procent av alla artiklar i ordboken, slumpmässigt utvalda.

Enligt detta test var mer än 98 % av de kontrollerade artiklarna korrekt indexerade. De två vanligaste felen visade sig vara dels att en eller flera momentsiffror ej blivit identifierade i betydelsebeskrivningen, dels att en uppslagsform tolkats som biform till det närmast föregående ordet trots att den faktiskt utgör ett nytt uppslagsord.

Det förekom också en del fel i artiklarnas huvud, biformer hittas inte alltid och inte heller ordklass. Blott i några enstaka fall hade indexeringen sparat ur helt och missat en större mängd moment eller uppslagsformer och det berodde då oftast på fel i indata som t.ex. obalanserade parenteser. Vi räknar med att snart ha en korrekthetsgrad på över 99 %.

Vi har även programmerat ett antal automatiska tester som bland annat letar efter ensamma moment, moment som inte kommer i ordning eller som bryter hierarkin bland momentnivåer. Vi söker även efter sammansättningar som ej kommer i alfabetisk ordning, vilket oftast är ett tecken på att något är fel i indexeringen.

3. Framtiden

Vi fortsätter arbetet med digitaliseringen av ordboken. Närmast på tur står att indexera häfte 37:1 som kom ut i december 2014 och som omfattar uppslagsorden UTSUG till VEDERSYN. Vi utför också återkommande stickprovskontroller för att hitta och åtgärda fel i indexeringen.

Nästa stora projekt blir att färdigställa en webbversion av ordboken som kan publiceras på nätet. En prototyp till en sådan webbsida finns redan idag och arbetet pågår med att utveckla denna till en färdig produkt.

The screenshot shows the Swedish Academy's online dictionary interface. At the top, there is a search bar with a 'Sök' button and navigation links for 'INDEX', 'FRITEXTSÖK', and 'HJÄLP'. The main content area displays the entry for 'FLOTTE sbst. 2' from 'Spalt : F-872 band 8, 1925'. The entry includes a detailed definition: 'FLOTTE *flot³ e²*, sbst.², l. (numera bl. starkt bygdemålsfärgat) FLÅTE *flå³ te²*, r. l. m.; best. **-en** ((⁺) **-flåttm RÅLAMB** 10: 42 (1691)); pl. **-ar**; äv. (numera bl. bygdemålsfärgat samt i Finl.) FLOTTA *flot³ a²*, sbst.², r. l. f.; best. **-an**; pl. **-or** ((⁺) **-er** (i vissa fall möjl. att hänföra till sg. **flotte**) 2*Krönl.* 2: 16 *Bib.* (1541), HALLENBERG *Hist.* 5: 260 (1796)).

Below the definition are sections for 'Formparentes' and 'Etymologi'. The etymology section contains the text: '[*flv. floti*, motsv. *d. flaade*, *isl. floti*, *feng. flota*, *eng. float*, bildat på det svaga rotstadiet till **FLYTA**; jfr **FLOTA**, sbst.¹, **FLOTT**, sbst.¹⁻², **FLOTT**, adj.¹⁻², **FLOTTA**, sbst.¹, **FLOTTA**, v.¹⁻²]

The main definition is followed by a numbered list item: '1) mängd stockar l. bjälkar l. dyl. som sammanfästs för att flyta på vattnet o. tjäna till (provisoriskt) fortskaffnings- o. transportmedel över vatten, stundom äv. avsedd för andra ändamål. *En i hast hoptimrad flotte. En flotte hopfogad av några stockar. Sammanfogade flottar av grenar och trädstammar. Pärlfiske från båt eller flotte. Fartyget var försett med så många livbåtar och flottar, att alla ombordvarande*

On the right side, there is a sidebar with a search filter set to 'former' and a list of related terms: flotte, ·flåte, ·flotta, A, flott-binda, flott-binda, flott-bindsle, flott-bom, flott-boplat, flott-bro, flott-brygga, flott-byggare, flott-båt, flott-förare, flott-lägga, flott-läggare, flott-lägnings-apparat, flott-lägningsdocka, flott-präm, flott-skeppare, flott-skuta, flott-stock.

Figur 7: Skärmlapp från indexeringsprojektets prototypsida.

Längre fram kommer vi, förutom att indexera nya häften efterhand som de kommer ut, att försöka fördjupa indexeringen. Det finns mycket som är intressant att identifiera, t.ex: hänvisningar inom och mellan artiklar, källor, brukligheter och fackområde, etymologier med mera.

Litteratur

Ordböcker

SAOB = *Ordbok över svenska språket utgiven av Svenska Akademien (Svenska Akademiens ordbok)*, 1893 ff. Lund.

SAOB i digitaliserad form (band 1–35) = <g3.spraakdata.gu.se/saob/> (november 2014).

Annan litteratur

Bäckerud, Erik (2014): Nydigitaliseringen av SAOB. I: Ruth Vatvedt Fjeld & Marit Hovdenak (red.): *Nordiske studier i leksikografi 12. Rapport fra Konferanse om leksikografi i Norden Oslo 13.-16. august 2013*. 95-105.

Lundbladh, Carl-Erik (1992): *Handledning till Svenska Akademiens Ordbok*. Stockholm: Norstedts.

OSA (1996) = Sture Allén, Yvonne Cederholm, Sofie Kokkinakis Johansson, Lena Rogström, Rudolf Rydstedt & Lars Svensson (1996): *Om svar anhålles. Rapport från projektet OSA*. GU-ISS-96-4 Research reports from the departement of Swedish, Göteborgs Universitet.

Erik Bäckerud
systemansvarig
Svenska Akademiens ordboksredaktion
Dalbyvägen 3
SE-224 60 Lund
erik.backerud@svenskaakademien.se

Ordnett, forlaget og brukerne

Øystein Eek

The article gives a presentation of the interaction between Kunnskapsforlaget and its customers in relation to the dictionary website <www.ordnett.no>. Ordnett was launched in 2004, and is the most widespread commercial dictionary website in Norway. Different tools have been developed to make contact between the publisher and the customers as easy as possible, and the customers give valuable feedback to both the editorial and technical development.

1. Ordnett og forlaget

Ordnett er en digital ordbok- og språktjeneste opprettet av Kunnskapsforlaget i 2004. Standardversjonen publiseres på nettstedet <www.ordnett.no>, i tillegg tilbys programvaren Ordnett Pluss, og en app-versjon for utvalgte ordbokbaser.

Porteføljen består av ca. 50 ordbøker, de fleste fra Kunnskapsforlaget, men omfatter også lisensierte titler fra Oxford University Press, PONS, Le Robert og (spansk) Larousse. Nettstedet omfatter en del hjelpetekster, for det meste hentet fra bokutgavene av de samme ordbøkene.

Tjenesten forvaltes i forlaget (2015) av en prosjektansvarlig redaktør med bistand av tre ordbokredaktører, dessuten av to IT-utviklere, tre salgs- og markedsføringsmedarbeidere og 1–2 kundebehandlere. Innholdet i ordbokbasene vedlikeholdes av et antall frilansleksikografer og forlagets egne redaktører.

2. Brukerne

På grunn av de mange flerbrukeravtalene er det umulig å vite antallet Ordnett-brukere, men trolig har ca. 300 000 – 400 000 tilgang gjennom egne abonnementsavtaler. De fleste Ordnett-brukere har tilgang gjennom utdannings- eller arbeidssted. Ordnett benyttes av et stort flertall av elevene i

videregående skole (14 av våre 19 fylkeskommuner pr. august 2015), dessuten i høyere utdanning, på biblioteker, og en rekke andre institusjoner. I tillegg er det noen tusen individuelle brukere.

3. Forretningsmodellen

Kunnskapsforlaget ANS er et privat firma som eies av de norske forlagene Aschehoug og Gyldendal. Innholdsutviklingen er fra starten av basert på forlagets portefølje av trykte ordbøker, men for titlene som utgis i både trykt og digitalt format, blir den mer og mer styrt av spesifikasjoner levert fra Ordnett.

Bortsett fra et forsøk med gratis norsk ordbok i startfasen, har Ordnett hele tiden vært en gjennomført betalingstjeneste, der brukerne betaler abonnement månedsvis eller for et år av gangen. Tilleggsinnhold som kryssordhjelp og minigrammatikker er gratis, forutsatt at man er registrert bruker. Noen av storkundeavtalene har kun ett års varighet, mens andre er flerårige. Gratis prøvetilgang gis i betydelig omfang, men for en kort tid, vanligvis et par uker. Appene kjøpes en gang for alle.

4. Bruerkommunikasjon før Ordnett

Ved forhandlersalg av trykte bøker er det minimal kontakt mellom forlag og brukere. Den trykte boken fremstår som et autoritativt produkt. Før Internettets tid måtte kunder med noe på hjertet skrive brev til forlaget, og dette forekom meget sjelden. Redaksjon og forfattere arbeidet i et tilnærmet vakuum. Anmeldelser av ordbøker i dags- og fagpresse forekom, men mange ordbøker har nok eksistert på markedet og vært solgt i titusener av eksemplarer uten å ha møtt noen form for offentlig interesse.

Omkring 1990 begynte Kunnskapsforlaget å utgi digitale ordbøker for installasjon på pc, først levert på disketter, deretter på CD-ROM, og senere på DVD. Kommunikasjonen med kundene/brukerne handlet først og fremst om tekniske spørsmål, knyttet til selve installasjonen og til kompatibilitet under ulike operativsystemer og maskintyper. Det ble ikke tilbudt oppgraderinger som fulgte utviklingen av operativsystemene, ei heller

utviklingen av de kontorstøtteverktøyene som ordbokprogrammene skulle samhandle med.

Salget til individuelle brukere skjedde først og fremst gjennom forhandler. Kommunikasjon med brukerne om innholdet i ordbøkene forutsatte fortsatt at man skrev et brev, senere eventuelt sendte en e-post. Svært få innspill nådde ordbokredaksjon og forfattere på denne tiden; situasjonen var dermed nokså lik den man hadde da det bare fantes trykte ordbøker. Bortsett fra et fåtall brev og andre meldinger fra høykompetente språkfolk, samt unntaksvis meldinger om rene trykkfeil, skjedde produktutviklingen helt og holdent på redaksjonens og forfatternes premisser.

5. Bruerkommunikasjon i dag

Med etableringen av Ordbnett i 2004 fikk vi en bredere og mer intens kontaktflate mot ordbokbrukerne, en mer differensiert brukerkommunikasjon, og en langt lavere terskel hos brukerne for å komme med innspill til forlaget. Muligheten for gratis testperiode gjør det enklere å gjøre seg kjent med et produkt, og på grunn av abonnementsmodellen er kundenes økonomiske risiko hvis man ikke blir fornøyd, lavere enn ved kjøp av en trykt ordbok. Konkurransen om kundenes oppmerksomhet er mye sterkere i og med at det finnes gratisprodukter – riktignok av varierende kvalitet. Markedet er blitt internasjonalisert, ettersom de samme nettproduktene er tilgjengelige over hele verden. For å imøtekomme behovene hos brukere som ikke har norsk som morsmål, har forlaget laget en parallellversjon av brukergrensesnittet i Ordbnett med engelsk tekst.

Det kan nok tenkes at respekten for autoriteten knyttet til forlagsproduktene er lavere enn før, og forventningene til at produktene skal imøtekomme den enkelte brukers personlige behov, er blitt sterkere. Om forventningene til respons på henvendelser er høyere, er vanskelig å vite, men det er ikke sikkert, siden vi ikke mottar så mange purringer på svar. Brukernes mer aktive holdning gir utgiverne langt større innsikt i hva brukerne er ute etter, og kan tilpasse produktutviklingen etter brukernes ønsker, hvis ressursene tillater det. Kundeinnspillene gjelder nok marginale, individuelle ønsker og behov, men gir samlet en pekepinn om hvor

forbedringspotensialet ligger. For eksempel dreier mange av kundeinnspillene til norskordbøkene seg om fagterminologi, noe som ikke er uventet i et samfunn med høyt utdanningsnivå.

5.1. Salg og markedsføring

Også salgsarbeidet innebærer kontakt med brukerne, på litt ulike vis etter salgsformen.

5.1.1. Anbud

I snart 10 år har salg til videregående skole i Norge hovedsakelig skjedd gjennom fylkeskommunale anbud. Kunden definerer sine behov gjennom en anbudsutlysning, og anbudet besvares gjennom et omfattende skriftlig tilbud, der beskrivelse av de enkelte produktene inngår. I forbindelse med at anbudet skal vurderes, blir det iblant arrangert møte mellom forlaget og representanter for fylkeskommunen, der produktportefølje, tekniske løsninger, servicetilbud, mulighet for kurs og annen kundeoppfølging blir diskutert. Tilsvarende prosesser gjennomføres for det offentlige markedet gjennom CRISStin-systemet. CRISStin – en institusjon under Kunnskapsdepartementet – inngår ettårig avtale på vegne av en rekke offentlige institusjoner, som så enkeltvis kan velge produktløsning. I senere tid er rammeavtaler for læremidler i noen grad benyttet også for kjøp av digitale ordbøker til videregående skole, istedenfor anbud.

5.1.2. Bedrifts- og institusjonssalg

Salg til bedrifter og institusjoner skjer hovedsakelig på forlagets initiativ ved personlig henvendelse fra salgavdelingen. Hvis kunden er interessert, prøver forlaget å sette sammen en relevant portefølje på tittelnivå. Kundene synes i det store og hele å akseptere innholdet og teknologien slik den er. Ønsker fra enkeltkunder om nye produkter kan sjelden imøtekommes, av økonomiske grunner, men det er en del kommunikasjon om enkeltartikler m.m. i ordbøkene. F.eks. utvides innholdet i Jusleksikonet fra utgave til utgave, ikke minst som følge av forlagets kommunikasjon med advokatkontorer og andre eksisterende og potensielle brukere på høyt nivå.

5.2. Sosiale medier

Ordnets Facebook-side har ca. 1 900 «Likes» pr. august 2015. Den redigeres av forlagets kundeservice, der studenter er ansatt på timebasis. Siden benyttes til å publisere språkstoff i vid forstand, ikke bare knyttet til forlagets ordinære utgivelser. Siden er først og fremst blitt et forum for forlaget selv, og har ikke etablert seg som et språklig diskusjonsforum. Det er lite av det som kommer inn på denne siden fra utenverdenen, som kan utnyttes av redaksjon eller produktutviklere.

5.3. Nyhetsbrev

Salgsavdelingen sender ut nyhetsbrev med e-post ca. 10 ganger i året, først og fremst for å informere om nye utgivelser, om større redaksjonelle oppdateringer og andre forbedringer, samt om spesialtilbud og andre salgskampanjer. E-postene går til både eksisterende og tidligere kunder. Hovedformålet med nyhetsbrevene er å vise kundene at forlaget stadig forbedrer produktporteføljen, samt å oppnå mersalg.

5.4. Teknisk support

Kunnskapsforlaget ønsker å fremstå som en leverandør med høyt servicenivå. Ettersom programmene er egenutviklet, har forlagets teknisk ansatte høy kompetanse og kraftig eierskap til løsningene, og prioriterer alltid teknisk support foran alt annet. Gjennom Ordnetts 11 år på markedet har det aldri vært stilt spørsmålsteget ved responstid og personlig oppfølging og service. Kappelstriden i markedet med Ordnetts hovedkonkurrent iFinger/Clarify har hatt som konsekvens at produktutviklingen må skje kontinuerlig. Kundeservice tar seg av førstelinjesupport, men kanalen inn til utviklerne er åpen hele tiden.

5.5. Kommunikasjon av betydning for produktutviklingen

Vi må her skille mellom kommunikasjon med innkjøpere og administratorer (f.eks. lærere) på den ene siden og enkeltbrukerne på den andre siden. Det har hendt at krav i anbud har slått direkte ned i produktutviklingen, f.eks.

ved lisensiering av enspråklige ordbøker for fremmede språk i tillegg til engelsk. Andre ganger har innspill kunnet utløse prosjekter, men disse vil ikke bli fullført så raskt som kunden kanskje ønsker. Etter at forlaget eventuelt har fått en anbudskontrakt, fortsetter kontakten med kunden på lavere nivå, for eksempel ved møter med lærere som skal bruke Ordnett. Kundemakten er meget stor i dette markedet, og forlaget satser mye på å imøtekomme realistiske ønsker om endringer. Slike endringer blir da gjort gjeldende for alle markeder, ettersom vi ikke har differensierte tilbud for ulike kundegrupper. En del av kundene innenfor profesjonsmarkedene identifiserer seg med yrkestitler o.l. når de sender kommentarer og ønsker, og disse innspillene blir behandlet med høy prioritet.

I 2014 åpnet Ordnett en forenklet løsning for å sende inn forslag til nye artikler inn til redaksjonen. Mange har benyttet seg av denne funksjonaliteten, og forlaget har fått mange hundre nyttige innspill. Mange innsendte forslag har utløst umiddelbar redaksjonell aktivitet, mens andre blir vurdert litt over tid. Det er viktig for oss å vise at kundenes tips tas alvorlig. Det gis automatiserte tilbakemeldinger til innsenderne, men intensjonen om personlige tilbakemeldinger har det ikke alltid vært mulig å oppfylle.

Kundenes innspill gjelder særlig språkene norsk og engelsk. Forslagene blir formidlet til den redaktøren som har ansvar for språket Ordnett-brukeren har hatt som aktivt språk i programmet da forslaget blir sendt inn. Frem til utgangen av oktober 2015 hadde norskredaksjonen registrert ca. 350 forslag som kan klassifiseres som seriøse. Noen av forslagene blir umiddelbart tatt inn i Tanum-basen, som har den største ordmassen, og så blir de senere vurdert for spesialordbøkene. På denne måten er ord som *ansvarliggjøre*, *handlingsreferat*, *bruksareal*, *arbeidsskade*, *innkvarteringsområde*, *selvforsterkende*, *utforskertrang* og *kulturfundamentalisme* kommet inn. De fleste forslagene gjelder sammensatte ord, som det jo finnes uendelig mange av, men vi sammenligner da med de sammensetningene som står fra før under samme førsteledd, og finner rett som det er at kundenes innspill bringer minst like aktuelle og utbredte ord inn i ordboken, som de vi allerede har. I en del tilfeller er forslagene også like relevante for engelskbasen som for norskbasen.

5.6. Kommunikasjon med redaksjonen som faglig ressurscenter

Kunnskapsforlaget får en del henvendelser om språkspørsmål som ikke er knyttet til Ordnett. Redaksjonen forsøker å svare etter beste evne, og selv om vårt fagmiljø ikke kan måle seg med Språkrådets, har vi tilgang på svært kvalifiserte medarbeidere. Vi legger vinn på å svare raskt og grundig, og har flere ganger fått hyggelige tilbakemeldinger fra innsenderne. Også slike henvendelser kan ha redaksjonelle konsekvenser. Som eksempel kan nevnes en fersk henvendelse om tvilstilfeller i bruken av bindestrek i sammensatte ord. Her ble svar fra en av spesialmedarbeiderne i språkverktøyet Språkvett umiddelbart formidlet til innsenderen. Også i andre tilfeller har vi benyttet ord- og håndbokforfatteres tid og kompetanse, og dette gir alltid gode resultater i kundekommunikasjonen.

6. Utnyttelse av søkestatistikk

Forlaget bruker ulike overvåkningsverktøy på ordnett.no, bl.a. en egenutviklet søkestatistikkmodul som er integrert i ordnett.no. Fra statistikkmodulene henter vi ut søkeord som gir 0 treff. Dette materialet gir innspill til nye artikler, og det gir informasjon om feil brukere gjør, som kan utnyttes til å bedre søkefunksjonaliteten.

7. Avsluttende bemerkning

Både redaksjonen og eksterne ordbokforfattere, og forlaget generelt, ser på den mer aktive dialogen med brukerne som en klar forbedring i forhold til den tidligere arbeidssituasjonen. Aktive brukere er en meget viktig ressurs for dagens ordbokutgivere. Brukernes aktivitet har også en økonomisk side, idet den stiller kostnadsfritt til forlagets disposisjon en mengde nyttig stoff. For forlaget, som skal leve av ordbokinntektene, kan de utvidede og forenklete kommunikasjonsmulighetene øke brukernes motivasjon for å benytte ordboktjenestene aktivt, og dermed motivere dem til å forlenge sine abonnementer. For forlaget er det viktig at de kundene som sender inn konkrete spørsmål, får svar raskt, og at svarene er godt begrunnet. Dette er en høyt prioritert oppgave for redaksjonen og dets nettverk. For redaksjonen

og forfatterne er innspillene om enkeltord nyttige i seg selv, men de gir også indikasjoner om hvilke emneområder de aktive brukerne er opptatt av, og dette kan gi støtet til mer omfattende forbedringsprosjekter. For eksempel vil en vennlig henvendelse fra en faglærer ved en av våre høyskoler resultere i et (riktignok ennå ikke gjennomført) prosjekt for gjennomgang av stikkordmaterialet innenfor kjemi, i norsk- og engelskbasene. Den tiden vi bruker på kundehenvendelser og -forslag, går nok ut over våre muligheter til å konsentrere oss om planlagte aktiviteter av mer generell karakter, som teknisk-leksikografiske dataforbedringer og større, systematiske innholdsrevisjoner. En fullstendig kundestyrt produktutvikling vil derfor ikke være vår strategi. Vi må i også årene fremover prøve å opprettholde en best mulig balanse mellom det planlagte og det improviserte. Ved å informere kundene om større, redaksjonelt initierte forbedringsprosjekter, i nyhetsbrev og på andre måter, håper vi at dette i sin tur vil øke kundenes årvåkenhet på disse områdene og stimulere dem til nye, nyttige innspill.

Øystein Eek
seniorredaktør, cand.philol.
Kunnskapsforlaget
Postboks 4432 Nydalen
NO-0403 Oslo
oystein.eek@kunnskapsforlaget.no

Facebook – et nyt brugersegment for Ordbog over Dansk Talesprog

Louise Hallstrøm Abildgaard, Amalie Glargaard & Carsten Hansen

Ordbog over Dansk Talesprog (ODT) – a Dictionary of Spoken Danish – is a project under the LANCHART Centre at the University of Copenhagen. Following ODT's online launch in October 2014 which received a lot of attention in Danish media, the ODT team decided to create a Facebook (FB) page for the project as well. This article describes the process of developing an explicit strategy for integrating the FB page and its audience into the work of the team, and how to differentiate between and manage two different platforms, namely the website and the FB page. Early on, the scope was to generate more traffic to the ODT website. However, due to FB having such a vast user group and being a different type of medium altogether, it became necessary to broaden that initial scope. Analyzing Facebook statistics, developing a strategy for publishing updates and particularly a decision to modify the very academic language of the website to better suit the FB format, all proved fruitful in generating FB traffic, if not traffic directly to the website. Thus, FB has become a separate channel of distribution, parallel to that of the website, but – at least potentially – reaching a much wider audience.

1. Indledning

Ordbog over Dansk Talesprog (ODT)¹ er et ordbogsprojekt ved Sprogforandringscentret på Københavns Universitet. Ordbogen er baseret på centrets talesprogs-korpus og ser sig selv som et talesprogsbaseret supplement til den etablerede, skriftsprogsbaserede leksikografi. Redaktionen arbejder for øjeblikket med et projekt om interjektioner – en ordklasse som er særligt forankret i talesproget. Hjemmeside og ordbog blev lanceret online i efteråret 2014, og i euforien efter den positive modtagelse i den faglige og almindelige offentlighed opstod ideen om også at lade

¹ Ordbog over Dansk Talesprog er tilgængelig på www.odt.hum.ku.dk

ordbogen repræsentere med sin egen side på Facebook. Det dobbelte brugersegment for hjemmeside og ordbog var klart defineret som (1) den professionelle sprogforsker og (2) den alment interesserede sprogbruger. Med facebookside stod vi pludselig over for et helt nyt og tredje brugersegment: facebookbrugeren.

I denne artikel redegør vi først for vores begrundelse for at lade ODT repræsentere på Facebook. Derefter rapporterer og reflekterer vi over vores tiltag for at øge facebooksidens fascinationskraft og dermed fastholde og øge antallet af følgere. Den sidste del af artiklen består i en statistisk rapport over aktivitet på facebookside samt en karakteristik af facebooksidens prototypiske følger. Formålet med denne øvelse er at undersøge om tiltagene for at øge facebooksidens fascinationskraft har virket.

2. Hvorfor Facebook?

Med et antal brugerprofiler der nærmer sig de to milliarder, er Facebook langt det største og mest udbredte sociale netværk på internettet. Alene af den grund er det interessant i forhold til markedsføring. Desuden er det i udgangspunktet gratis, og brugerne kan gøre meget for at fremme deres sag også uden at afsætte midler i budgettet.

I sin natur er en facebookside hurtigere at bruge og redigere end en traditionel hjemmeside. Mekanismerne omkring *følgere*, *likes*, *delinger* og *kommentarer* giver en mere direkte og umiddelbar berøringsflade med brugerne, og netværkskarakteren giver mulighed for at udnytte en virtuel centrifugalkraft til potentielt at nå langt bredere ud end den umiddelbare målgruppe. Desuden er sproget på de online sociale netværk tydeligt præget af selve måden disse netværk fungerer på hvilket f.eks. bevirker en meget høj grad af talesprogsnære udtryksvarianter. Koblingen mellem et talesprogsnært medie og talesproget som studieobjekt ligger lige for.

Til trods for at målet med og effekten af en side på Facebook kan forekomme diffus, navigerer ODT-redaktionen efter nogle erklærede pejlemærker.

2.1. Formål og målgruppe

Der foreligger ingen indgående analyser af trafikken på ODT's hjemmeside. Alligevel har redaktionen et klart billede af den gennemsnitlige besøgende som en midaldrende person der færdes i eller omkring et akademisk miljø. Ønsket med facebookside er at nå ud til et nyt, bredere og også yngre brugersegment, herunder studerende. Desuden er det tanken at udnytte facebookside til at nå ud til folkeskole-, gymnasie- og sprogskolere der vil kunne drage nytte af både hjemmesiden, men især også af facebookside med den mere lette indgang til talesprogsordbogen i deres undervisning.

2.2. Fordele ved Facebook

Facebook har den fordel at konceptet gør det lettere at nå ud til folk gennem diverse facebookfora, samt at nyheder på Facebook kan gå viralt for dermed at sprede sig som ringe i vandet. Herudover giver facebookside også mulighed for at holde ordbogsbrugere og ordbogsinteresserede opdaterede om nye artikler og nye tiltag fra ODT, ligesom den giver mulighed for at interagere med facebookfølgerne der eksempelvis kan inspirere til nye lemmer.

2.3. Udfordringer ved Facebook

Et nyt medie giver også nye udfordringer. Først og fremmest er en facebookside tidskrævende idet følgerne forventer jævnligt at blive opdateret om nyheder fra ODT. Herudover bliver Facebook en ny medieplatform som følgerne kan henvende sig på med spørgsmål og andet der skal tages stilling til og svares på. Endelig er der en risiko for at udelukke ODT-interesserede der ikke benytter sig af Facebook.

2.4. Hjemmeside vs. facebookside

Facebookside skal ses som en nem måde for brugere at finde information om ODT, men samtidig kan man frygte at facebookside fjerner opmærksomheden fra hjemmesiden. Det har vi forsøgt at løse ved at linke til

hjemmesiden via facebookside og på den måde kæde de to medieplatforme sammen.

Koblingen mellem de to medieplatforme giver en ny udfordring idet hjemmesiden og facebookside har forskelligt udtryk – særligt sprogligt. På hjemmesiden anvendes et fagligt sprog² som henvender sig til den sprogvidenskabeligt og alment interesserede bruger, mens facebookside anvender et mere talesprogsnært og humoristisk sprog. Der opstår et markant registerskift når facebookbrugerne klikker på hjemmesidelinket.

3. Fra spontanitet til strategi

Beslutningen om at etablere en facebookside for ODT³ blev vedtaget spontant oven på en overvældende medierespons efter lanceringen af hjemmeside og ordbog. De første facebookopslag gav god respons men blev hurtigt ensformige og gav stadig færre likes, kommentarer og delinger.

Der måtte altså formuleres en strategi for at fastholde følgernes interesse, og overvejelserne var primært:

- Hvordan får man fat i målgruppen?
- Hvad skal forholdet være mellem hjemmeside og facebookside?
- Hvilke typer opslag skal man lave for at fange følgernes opmærksomhed?
- Hvor tit skal der laves opdateringer, og hvornår på døgnet vil man bedst kunne fange brugerne?

Som svar på disse spørgsmål enedes redaktionen om en række tiltag som en del af en samlet handlingsorienteret strategi.

Helt overordnet blev det en hovedpræmis for strategien at se facebookside som *et selvstændigt parallelspor til hjemmesiden*. Det anvendte sprog på facebookside adskiller sig som nævnt fra sproget på hjemmesiden. I stedet for at homogenisere de to medieplatforme blev det således besluttet at facebookside skal fungere som selvstændig indgang til viden om

² Se Carsten Hansen (2015) vedr. beskrivelsessproget i ODT.

³ ODT's facebookside er tilgængelig på:
<www.facebook.com/ordbogoverdansktalesprog>.

talesproget da facebookbrugeren bliver et nyt brugersegment for ODT – et brugersegment som der skal appelleres til på en anden måde end til hjemmesidebrugeren. Ud fra denne præmis formulerede redaktionen følgende konkrete tiltag for udformning af opslag på facebookside:

1) *Multimodal kommunikation*

For at komme uden om det hidtil noget ensformige udtryk i facebookopslagene blev det besluttet at benytte varierende illustrationsmateriale i form af fotos, tegninger, video eller lyd som illustration af de enkelte opslag.

2) *Omskrivninger af artikler til talesprogsnært sprog*

For at bevare facebooksidens parallel-identitet blev det besluttet at lægge ordbogsartikler direkte ud på Facebook men i et andet og mere talesprogsnært sprog end på hjemmesiden. På denne måde gør det ikke noget hvis brugerne ikke kigger på hjemmesiden – der leveres viden direkte på Facebook.

3) *Inddragelse af brugerne*

Den dalende interaktion omkring facebookopslagene kaldte på en decideret invitation til brugerne om at deltage. Derfor blev det besluttet at redaktionen kan henvende sig til brugerne og f.eks. spørge hvad de synes skal være det næste lemma i ODT.

4) *Ramme brugerne når de er online*

Facebook er et medie hvor man konstant bliver oversvømmet med nyheder. Opdateringer skal derfor være spektakulære og fængende så de ikke drukner i nyhedsstrømmen, ligesom de skal lægges ud på det rigtige tidspunkt for at brugerne ser dem. Det blev derfor vedtaget at lægge opslagene ud på hverdage omkring kl. 15 da undersøgelser har vist at der er mest aktivitet på Facebook på dette tidspunkt. Derudover blev det besluttet at det ville være nok med gennemsnitligt én statusopdatering om ugen – det er en hårfin balance at bibeholde brugernes interesse uden at trætte dem. For mange opslag kan potentielt irritere brugerne med det resultat at de stopper med at følge siden. Derfor vurderedes en ugentlig opdatering som optimal.

3.1. Strategimål

Ved implementering af ovenstående strategi var det konkrete mål at opnå minimum 500 følgere på facebookside samt at hvert enkelt opslag skulle have minimum 10 likes/delinger.

3.2. Et eksempel fra ODT

Et enkelt eksempel på hvordan præsentationen af et nyt lemma er bearbejdet kraftigt på vejen fra hjemmeside til Facebook, er *fuck*. Artiklen i selve ODT, dvs. på hjemmesiden, ser omtrent sådan ud:

Fuck

[fʌg], [fʌ:g], [f:ʌg], sjældnere [fwʌg] (symbolforklaring)

Bruges til at forstærke talers eget udsagn eller som en emotiv kommentar på talerens egen eller samtalepartnerens ytring. **Fuck** kan også bruges som en kommentar på fysiske begivenheder.

1: Forstærkende funktion. **Fuck** kan både forekomme som det første ord i en ytring og dermed forstærke talerens efterfølgende udsagn, det kan forekomme midt i en ytring og dermed forstærke talerens igangværende udsagn, og det kan forekomme lige efter en ytring og dermed forstærke talerens foregående udsagn.

1.1: Forstærkende funktion i forbindelse med adverbiet *hvor*

2: [Emotiv](#) funktion

2.1: Udtrykker [negativ attitude](#) i forbindelse med for eksempel ærgrelse eller utilfredshed

2.2: Udtrykker [positiv attitude](#) i forbindelse med for eksempel overraskelse eller imponerethed.

2.3: Medlevende kommentar på fortalt begivenhed. Taleren sætter sig i den anden talers sted og udbryder **fuck** på vegne af samtalepartneren. I denne betydning danner **fuck** ofte en ytring i sig selv.

3: Kommenter på [fysisk begivenhed](#). Kommentaren er henvendt til taleren selv og skal ikke høres eller kommenteres af andre samtaledeltagere. **Fuck** bliver derfor ofte sagt hviskende og danner ytring alene

På Facebook er præsentationen noget anderledes (opslag på ODT's facebookside den 20. april 2015):

Fuck hvor sygt nice med en fuck-definition!
'Fuck' bruges til at:
1) understrege ting
2) udtrykke utilfredshed
3) udtrykke at man er imponeret eller overrasket
4) vise medfølelse
5) kommentere på noget der sker i rummet

Som det tydeligt fremgår, har facebookgengivelsen løsrevet sig meget fra selve ODT-artiklen. Dels præsenteres på Facebook kun den helt overordnede essens af artiklens mere komplekse analyse, dels etablerer Facebook-opslaget sin egen organisering af stoffet ved ikke at bibeholde hierarkiet og niveauforskellene fra artiklen, men fx etablerer sin egen nummerering. (Bemærk at der af praktiske grunde ikke er tale om en 1:1-gengivelse af hverken hjemmesidens eller FB's præsentation, men blot af selve teksten.)

4. Resultater af den implementerede strategi

Efter i et par måneder at have arbejdet ud fra strategien tegner der sig et af-læseligt resultat af dens effekt. Delmålet på minimum 500 følgere er nået, og ODT's facebookside har i skrivende stund 654 af slagsen. Derudover er det gennemsnitlige antal likes per opslag over 10 – faktisk ligger det på knap 15 hvor der før var et gennemsnit på godt 6 likes. Dette er illustreret i Figur 1.

Overordnet ser det altså godt ud. For at få et mere detaljeret billede af strategien fremlægges nu de enkelte elementer for at vise hvordan det er gået med at fastholde strategielementer samt evt. effekten af dem.

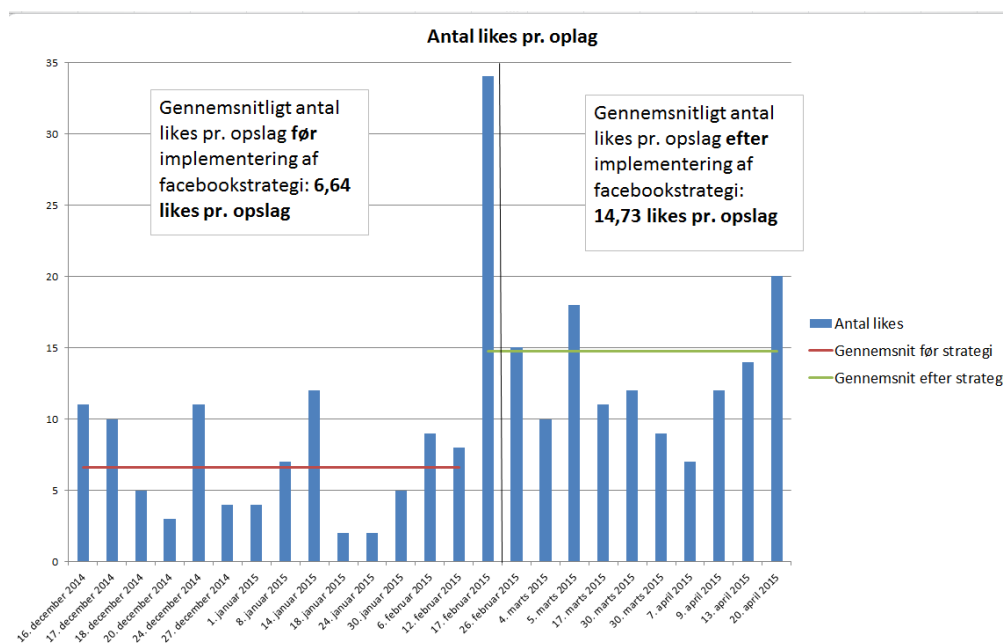
4.1. Multimodal kommunikation

Inddragelse af billeder og lyd har fungeret særdeles godt. Da ODT den 9. april lagde et lydclip op på facebookside, gav det ikke det største antal

likes, men til gengæld fik facebooksideen syv nye følgere – det højeste antal på så kort tid siden debuten på Facebook. Man kan derfor se at videoen har spredt sig og skabt opmærksomhed om facebooksideen.

Da målet om 500 følgere blev nået den 17. februar, blev der uploadet et billede af de tre redaktører der sagde tak til følgerne. Billedet er det opslag der har givet flest likes overhovedet. Det tyder altså på at følgerne også er glade for at få sat ansigt på projektet.

ODT har også benyttet sig af at linke til andre der har talt om interjektioner og talesprog generelt. Bl.a. er der blevet linket til et klip fra tv-programmet Monte Carlo på DR3 der laver deres egen definition af *hovs* funktioner. Det reagerede facebookfølgere positivt på og kvitterede med 17 likes og to kommentarer.



Figur 1: Antal likes per opslag før og efter strategi.

4.2. Omskrivning af artikler til lettilgængeligt sprog

Også de omskrevne artikler skrevet i et lettere og mere facebookvenligt sprog ser ud til at være populære. En omskrivning af artiklen om *fuck* fik 21

likes, og omskrivningerne af *aha* og *hold (da) op* fik hver 11 likes. *Hold (da) op*-omskrivningen blev desuden delt på en side der henvender sig til spansktalende dansklørnere. Det demonstrerer at facebookside fungerer efter hensigten og bliver brugt af den ønskede målgruppe, nemlig et nyt, bredere og også yngre brugersegment, herunder studerende, jf. ovenfor.

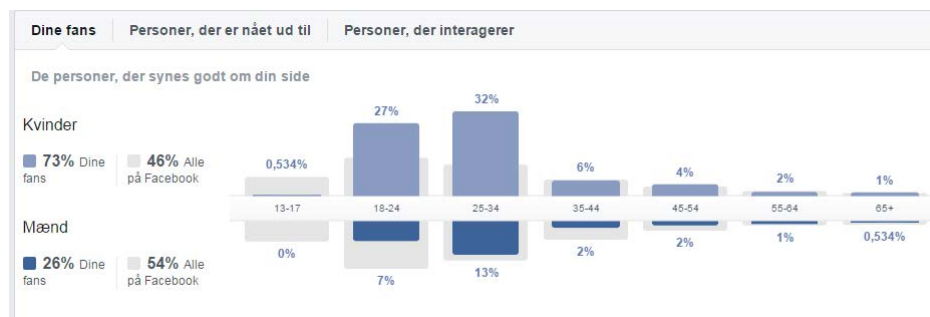
4.3. Inddragelse af brugerne

Indtil videre har der ikke været den store succes med at inddrage følgerne i interaktion på facebookside. I et opslag blev følgerne opfordret til at komme med forslag til nye lemmakandidater til ordbogen, men ud af de på det tidspunkt 541 følgere kom kun tre med forslag – langt færre end håbet. Det kunne tyde på at denne form for brugerinddragelse ikke henvender sig til målgruppen, og at man i stedet skal forsøge sig med andre initiativer. Desuden lærte vi en kommunikativ lektie af initiativet: Facebookredaktøren skyndte sig nemlig at takke de få der havde responderet, og mod hensigten kan det være blevet opfattet som en lukning af muligheden for at komme med flere forslag.

4.4. Ramme brugerne når de er online

Foruden nogle få opslag der har været nødvendige at poste på et andet tidspunkt end omkring kl. 15 (f.eks. nogle enkelte morgenopslag da projektleder Carsten Hansen skulle medvirke i radioprogrammet Sproglaboratoriet om formiddagen), har denne strategi været fastholdt, og opslag lægges ud ved 15-tiden en gang om ugen. Det fungerer godt.

Som det ses af Figur 2 nedenfor, er vores prototypiske facebookfølger en kvinde mellem 18 og 34 år – hele 59 % af ODT's følgere tilhører denne kategori. Det er dog også hovedsageligt denne gruppe der har en facebookprofil, og derfor er det ikke underligt at følgerne primært falder inden for denne gruppe. De yngre mellem 13 og 17 rammes stort set ikke (kun 0,5 % af ODT's følgere er piger i denne alder), og de ældre aldersgrupper rammes kun til dels.



Figur 2: ODT's facebookfølgere, alder og køn.

Med hensyn til geografisk placering ses det af Tabel 1 at langt hovedparten af ODT's følgere kommer fra København og omegn (304 personer). Resten kommer hovedsageligt fra Aarhus (71 personer), og herefter er følgerne fordelt med 5-6 stykker i hhv. Odense, Roskilde, Helsinki og Reykjavik samt i andre byer med mindre end fem følgere – de er ikke vist i tabellen nedenfor. Det er værd at bemærke at følgerne altså synes centreret i de to største byer i Danmark, som desuden er de to største universitetsbyer, ligesom Odense og Roskilde, som er de næste på listen, er universitetsbyer. Man fristes til at udlede at ODT især når ud til facebookbrugere fra det akademiske miljø i universitetsbyerne.

By	Antal følgere
København og omegn	304
Aarhus	71
Odense	6
Roskilde	6
Helsinki	6
Reykjavik	5

Tabel 1: ODT's facebookfølgere, geografisk placering.

Ud fra ovenstående kan det konkluderes at facebook siden til en vis grad når ud til den ønskede målgruppe: unge studerende. Det er svært at se om den

når ud til andre end vi ville være nået ud til med hjemmesiden, da der ikke foreligger andre tilgængelige sociokulturelle oplysninger om følgerne end køn, alder og geografisk placering.

5. Hvem er ODT's facebookfølgere

Programmeringen bag Facebook giver god mulighed for at trække brugerstatistik ud af systemet. Det er således forholdsvis uproblematisk at efterprøve hvorvidt de opstillede mål med hensyn til en udvidet brugergruppe er nået

6. Opsummering

Facebook er en ny platform for ODT som sikrer at ordbogen når ud til et nyt og bredere brugersegment. Det er et tidskrævende projekt som har krævet strategiovervejelser i forhold til målgruppe, typer af opslag og hensigtsmæssig distribution af opslagene.

Derfor har vi i ODT valgt at fokusere på (1) at være multimodale, (2) at kommunikere med vores brugere i et sprog som kan favne flere og andre end det lukkede akademiske brugersegment, (3) at appellere direkte til brugerne – samt (4) at kommunikere med brugerne i de tidsrum og det omfang som formodes at være passende.

Alt i alt må ODT's facebookside og redaktionens facebookstrategi siges at have været en succes. Facebooksiden har opnået langt over 500 likes (p.t. 654), og vores opslag har, siden strategien blev implementeret, gennemsnitlig fået over 10 likes hver. Herudover er der kommet mere variation i opslagene uden at fjerne fokus fra det primære formål: at promovere Ordbog over Dansk Talesprog og alle dens mange muligheder.

7. Refleksioner

Facebook er et markedsføringsinstrument hvis effekt skal tages seriøst. Som det ser ud nu, benytter mange små og store virksomheder og organisationer Facebook som talerør i bestræbelserne på at optimere deres

gennemslagskraft. Spørgsmålet er dog hvor længe denne trend holder: De sociale medier og deres rækkevidde ændrer sig hele tiden, og man må som organisation være forberedt på at skulle stifte bekendtskab med flere og nye medieplatforme hvis man vil holde fast i sin målgruppe og ramme nye. Hos os i ODT har vi i første omgang tilsluttet os Facebook som et særdeles veletableret og vidtfavnende socialt medie, og vi følger spændt med i udviklingen – for hvem ved om Facebook har samme pondus om et par år?

Litteratur

Hansen, Carsten (2015): Beskrivelsessproget i *Ordbog over Dansk Talesprog*. I: *LexicoNordica* 22, 57-75.

ODT's hjemmeside: <odt.hum.ku.dk/>.

ODT's facebookside: <www.facebook.com/ordbogoverdansktalesprog>.

Louise Hallstrøm Abildgaard
redaktør ved ODT
Sprogforandringscentret
Københavns Universitet
Njalsgade 136
DK-2300 København S
cjsx999@alumni.ku.dk

Amalie Glargaard
redaktør ved ODT
Sprogforandringscentret
Københavns Universitet
Njalsgade 136
DK-2300 København S
zhc736@hum.ku.dk

Carsten Hansen
lektor, projektleder
Sprogforandringscentret
Københavns Universitet
Njalsgade 136
DK-2300 København S
carhan@hum.ku.dk

Hvordan har netbrugernes præference for forskellige emnespecifikke ordbøger ændret sig i perioden 2006-2014?

Poul Hansen

The demand for 38 bilingual Danish/Swedish online dictionaries in the period 2006-2014 was statistically investigated. The results show that demand often varies with characteristic patterns in time and space. However, there are indications of a downward trend in user take-up and engagement. In future, more methods may be required to reach out to the users.

1. Indledning

Langt de fleste mennesker benytter i dag internettet som et redskab til at søge informationer og online-ydelser af forskellig art. At følge og undersøge hvordan disse brugere forholder sig til og finder frem til nettets ordbøger er af stor interesse for leksikografer (jf. Lorentzen & Theilgaard 2012).

I denne artikel undersøger jeg de langsigtede forandringer, der er indtruffet i forbindelse med benyttelsen af 38 bilingvale dansk/svenske net-ordbøger i en 9-årsperiode i hhv. Danmark og Sverige.

De fire hovedspørgsmål, som jeg har stillet, er:

- Hvilke typer af ordbøger har været mest populære?
- Hvordan ser forandringen over tid ud?
- Er der forskelle mellem danske og svenske brugeres adfærd?
- Er der forskelle mellem private og professionelle brugeres adfærd?

2. Materiale

Materialet i dette studie blev først udviklet som et støttemateriale til brug ved kvalificeret oversættelse mellem dansk og svensk og bestod oprindeligt af 15 tosprogede specialordbøger med ca. 30.000 opslagsord i hver sprog-

retning. I 2005 blev alt lagt ud på nettet, og det udvikledes i løbet af fire år til at omfatte 38 tematiske specialordbøger med 54.000 opslagsord, som forelå både i en svensk-dansk version for danske brugere og i en dansk-svensk version for svenske brugere (se Tabel 1 for en oversigt over ordbøgerne og deres temaer).

De enkelte ordbøger er blevet lagt ud som selvstændigt profilerede under-sider, som ganske vist indgår i en overordnet klikbar og søgbar struktur på hjemmesiden, men som får 80-90 % af trafikken direkte fra Google, når ordbogsbrugere søger ord på nettet via Googles søgemaskine. Ordbøgerne placerer sig ofte i toppen af de organiske ("almindelige") søgeresultater på grund af en forholdsvis høj PR (Page Rank) for den samlede hjemmeside.

Besøgsantallet for begge sider har ligget på et relativt konstant niveau og har efter en kort opstartsperiode varieret mellem 15.000 og 25.000 unikke besøg om måneden. I den sidste del af perioden var der ingen signifikant forskel i antallet af besøg på den danske og svenske side.

Materialet foreligger med svensk-dansk sprogetning på den danske side <www.dansk-og-svensk.dk> og med dansk-svensk sprogetning på den svenske side <www.danska-svenska.se>. Begge sider har PR 5.

Ordbogsprojektet modtog støtte af Nordplus i 2008, og ordbøgerne er blevet beskrevet og refereret i litteraturen (Pálfi 2011, Kristensen 2012).

3. Metode

I undersøgelsen af brugernes adfærd blev der benyttet to redskaber, 1) statistisk analyse af efterspørgslen efter ordbøgerne på serveren og 2) statistisk analyse af mailede spørgsmål fra ordbogsbrugere.

Den statistiske analyse af efterspørgslen blev baseret på variablen "hits", som nærmere bestemt er det antal gange en vis side (ordbog) efterspørges på serveren som følge af søgeaktiviteter på nettet eller trafik via direkte links.

Analysen af de indkomne spørgsmål bestod i at strukturere og beskrive indholdet i ca. 1000 e-mails, som blev modtaget i tidsrummet 2008-2014 fra de brugere, der havde søgt forgæves i ordbøgerne. Data var de angivelser, som brugerne uopfordret anførte i deres henvendelser, f.eks. spørgeomne, titler, virksomhedsnavne, adresser etc.

Beregninger og diagrammer blev udført i programmet MS Excel.

4. Resultater

4.1. Ordbøgernes relative popularitet

Med hensyn til ordbøgernes indbyrdes popularitet fremgår det af Tabel 1, at 21 ud af de 38 ordbøger var blandt de 30 mest besøgte sider i undersøgelsesperioden. Der var 17 ordbøger, som ikke opnåede dette popularitetsniveau, herunder de relativt omfattende artsordbøger.

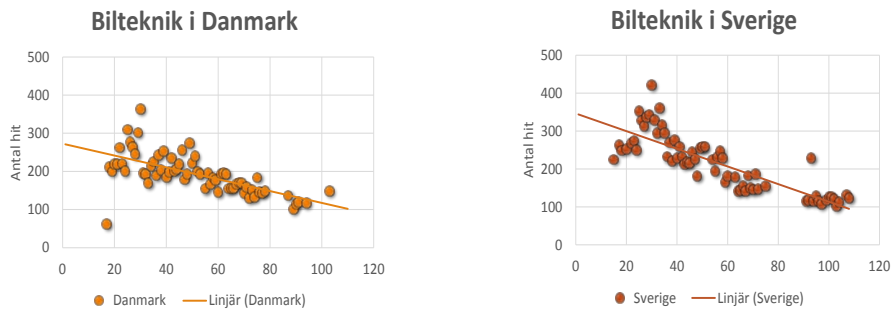
Ordbog/Tema	DK	S	Ordbog/Tema	DK	S
”TEKNIK”			”HVERDAGSORD”		
Data (4200)	29	16	Hverdagsord (7000)	104	95
Bilteknik (1500)	67	68	Ordpar (2500)		
Sikkerhed (1000)			Forkortelser (1100)	83	79
Byggeteknik (850)	54	95	Falske venner (950)	13	78
Håndværktøj (700)	35	49	Modsat gramm. køn (700)		
Arbejdskøretøjer (700)			Småord (80)	53	
Plasmaskæring (500)			”SLANG”		
MS*-kommandoer (275)			Nydanske ord (900)		
Svejsning (250)		3	Dansk chatsprog (430)	84	82
”FAGORD”			Skældsord (300)	84	
EU (8100)	49	76	Svensk chatsprog (270)	78	81
Fødevarer (1000)	53	94	Kælenavne, personer (220)	70	
Medicin (900)	100	94	Multietnisk sprog (200)		
Landbrug (750)	18	22	Bandeord (200)	85	2
Socialt arbejde (750)	22	78	Kælenavne, byer (100)		
Beklædning (500)	13	52	”NATUR”		
Kontor (400)			Fugle (5800)		
Lystfiskeri (400)			Planter (5000)		
Tilsætningsstoffer (320)			Svampe/Laver (3500)		
Kemikalier (250)			Naturtyperarter (400)		
			Botaniske termer (400)	15	26

Tabel 1: Den relative popularitet for undersøgelsens 38 ordbøger angivet som det antal månedlige gange i perioden 2006-2014, som ordbøgerne var blandt de 30 mest efterspurgte undersider på hhv. den danske (DK) og den svenske (S) hjemmeside. Antal lemma er angivet i parentes. * MS=Microsoft.

4.2. Tidsmæssige udviklingstendenser

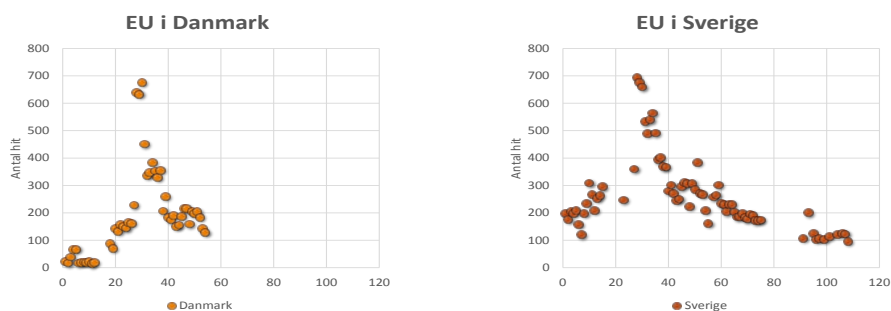
Som eksempel på den tidsmæssige variation præsenteres i de følgende diagrammer for ordbøgerne *Bilteknik*, *EU*, *Fødevarer*, *Falske venner* og *Bandeord*, hvor antal hits på den danske og den svenske side er gengivet for hver måned i perioden 2006-2014.

De biltekniske ordbøger i Danmark og Sverige havde en relativt overensstemmende tidsvariation med en svagt, nærmest lineært faldende tendens i en stor del af perioden, se Figur 1. En kort periode i opstarten var kendetegnet af vækst.



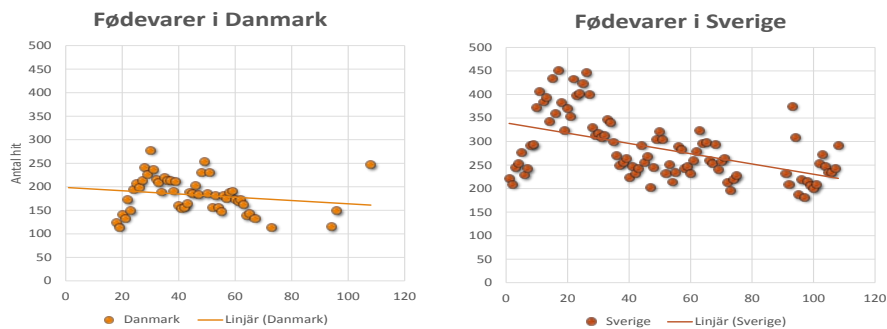
Figur 1: Tidsvariation for *Bilteknisk* ordbog i hhv. Danmark og Sverige, med antal hits som en funktion af antal måneder.

En noget anderledes kurve var gældende for *EU*-ordbogen, se Figur 2. Her var der en skarpt markeret top i begyndelsen af 2009 (måned 30) både på den danske og den svenske side med et tilsyneladende brat fald derefter.



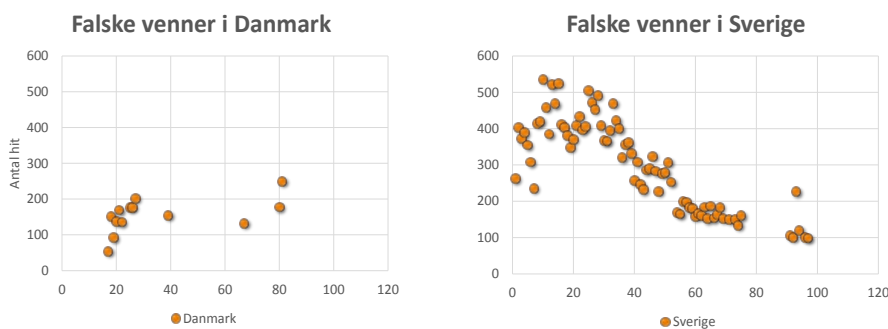
Figur 2: Tidsvariation for *EU*-ordbogen i hhv. Danmark og Sverige, med antal hits som en funktion af antal måneder.

Et noget mere komplekst forhold blev indikeret for *Fødevarer*-ordbogen, se Figur 3. Her var der niveauforskell mellem den danske og den svenske ordbog med hensyn til antal hits. Samtidig kendetegnes begge diagrammer af noget, der ligner fluktuationer. Begge diagrammer har dog generelt en fallende tendens.



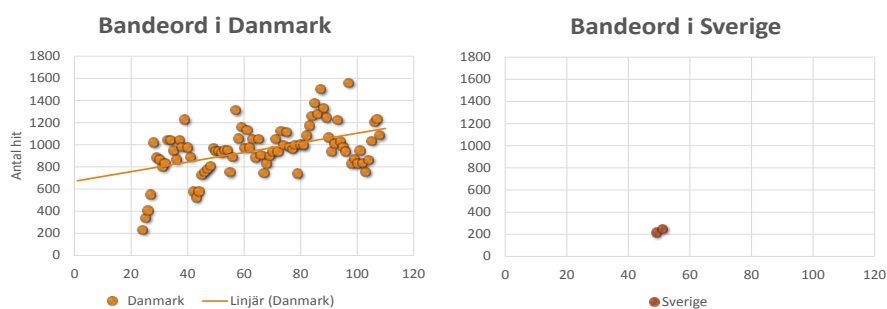
Figur 3: Tidsvariation for *Fødevarer*-ordbogen i hhv. Danmark og Sverige, med antal hits som en funktion af antal måneder.

Udviklingen for *Falske venner*-ordbogen adskiller sig betydeligt mellem Danmark og Sverige, se Figur 4. Mens tendensen er sporadisk og utydelig i det danske diagram med et relativt lavt og aftagende antal målepunkter, er der tale om en tydelig og mere klar tendens i det svenske diagram. Desuden er der et generelt højere popularitetsniveau for ordbogen i Sverige i det meste af perioden.



Figur 4: Tidsvariation for *Falske venner*-ordbogen i hhv. Danmark og Sverige, med antal hits som en funktion af antal måneder.

Det omvendte forhold gør sig gældende for diagrammerne for *Bandeord*-ordbogen, se Figur 4. Ordbogen kendetegnes af høj og stigende popularitet i Danmark, mens der kun er to målepunkter for popularitet i Sverige. Det danske diagram kendetegnes af store variationer i perioden, men en generelt stigende tendens kan anses for at være signifikant.

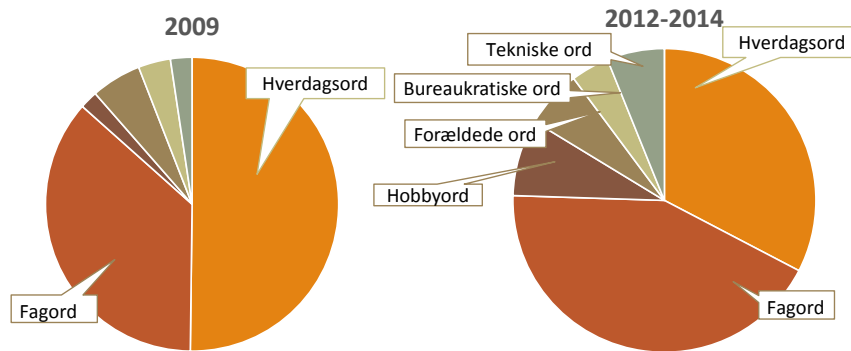


Figur 5: Tidsvariation for *Bandeord*-ordbogen i hhv. Danmark og Sverige, med antal hits som en funktion af antal måneder.

4.3. Analyse af brugernes spørgsmål

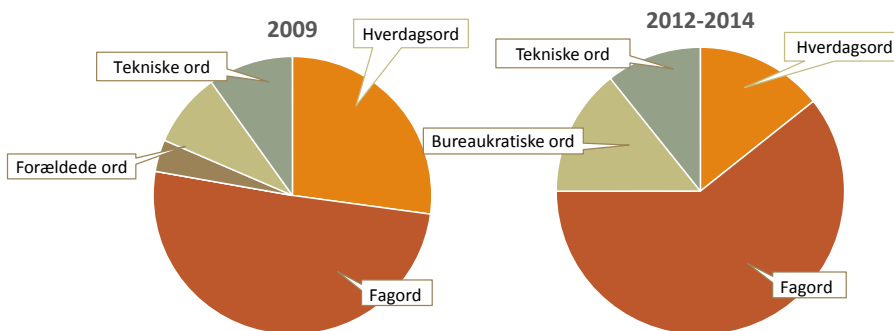
I perioden 2008-2014 blev der som tidligere nævnt modtaget ca. 1000 mails med spørgsmål fra ordbogsbrugere, som ikke fandt det ord eller den forklaring, de ledte efter. Analysen viser en kraftigt aftagende spørgefrekvens, fra ca. 30 pr. måned i 2008 til ca. 1 pr. måned i 2014.

I Figur 6 præsenteres spørgekategorierne for de private brugere. Her ses generelt et fald i andelen af spørgsmål vedrørende ”hverdagsord” og en tilsvarende stigning i andelen af ”fagord” og ”specialord”. Især ”hobbyord” og ”tekniske ord” fik en mere fremtrædende rolle i 2012-2014.



Figur 6: De relative spørgekategorier for private brugere i 2009 og 2012-2014.

For de professionelle brugere blev der noteret en lignende forandring, men med et andet udgangspunkt, se Figur 7. Her var andelen af "hverdagsord" nemlig lav allerede fra begyndelsen i 2009, men aftog derefter yderligere i 2012-2014. "Fagord" var den dominerende type og de udgjorde en endnu større andel i 2012-2014. En anden type ord, som også fik større andel, var "bureaukratiske ord".



Figur 7: De relative spørgekategorier for professionelle brugere i 2009 og 2012-2014.

5. Diskussion

5.1. Ordbøgernes indbyrdes popularitet

Der er stor forskel mellem de studerede ordbøgers popularitet. Ikke uventet har de ennemæssigt brede og ordrige ordbøger ofte størst popularitet. Nogle ordbøger skiller sig imidlertid ud ved at være efterspurgt trods et lille antal ord. Omvendt er der eksempler på ordbøger, som er relativt omfattende, men mindre efterspurgt, f.eks. artsordbøgerne *Fugle* og *Planter*.

Det er naturligvis ikke muligt at drage nogle sikre konklusioner omkring dette. Der er så mange ting, der kan spille ind, f.eks. målgruppernes størrelse, informationsbehov og deres medieadfærd. Men et punkt, som kan diskuteres, er, i hvilken udstrækning popularitetsforskellene skyldes ordbøgernes forskellige grad af unikhed på nettet. Artsnavne vil for det meste kunne findes andetsteds på nettet og med mere omfattende information. Derimod er fortegningerne over f.eks. bandeord og skældsord i temaet "Slang" relativt unikke på nettet og har en ekstra værdi for mange brugere, fordi de giver en oversigt og en klassificering af udtrykkene. Samtidigt er det åbenbart, at bandeordslisten og øvrige ordbøger inden for "Slang"-temaet opfattes som både kuriøse og underholdende. De har været omtalt i en række sociale medier (f.eks. på Facebook og Google Groups).

5.2. Forklaringer til den tidsmæssige udvikling

Mange af ordbøgerne i dette projekt har oplevet faldende trafik, hvilket kan have forskellige forklaringer. En forklaring kan naturligvis søges i sidernes udformning og de ulemper, som dette medfører (jf. Kristensen 2012), men også relativt populære netordbøger melder i dag om tilbagegang (jf. Stora-Ordlistan.se på <www.sis-index.se>). Man bør ikke se bort fra, at der kan være mange andre årsager end selve udformningen.

Først og fremmest kan der peges på, at antallet af aktører er steget kraftigt. Mange etablerede virksomheder og institutioner med stor gennemslagskraft har reorganiseret sig i forhold til internettets muligheder og gjort deres resurser tilgængelige og populære. De i dag vigtigste resurser for ordbogsbrugere, f.eks. Google Translate, fandtes ikke på nettet i undersøgelses-

periodens første år. Dertil kommer, at Googles søgeresultater i dag i høj grad er præget af, at mange aktører mere eller mindre forsøger at købe sig til placering via linksamarbejde, søgemaskineoptimeringer og sponsorannoncer. Flere og flere søgeord er omfattet af optimeringsbestræbelser og styring til kommercielle sider. Sider med lav informationsværdi (lav PR-værdi) placerer sig stadig oftere foran andre, mere videnstunge sider i de organiske søgeresultater.

EU-ordbogen (Figur 2) er et eksempel på tab af brugere til ”konkurrenter”. Det bratte fald i kurven kan ses i sammenhæng med, at EU et stykke inde i undersøgelsesperioden gjorde en lang række omfattende terminologi- og oversættelsesressourcer tilgængelige online.

På brugerniveau er der også sket store forandringer. Brugere er blevet dygtigere til at håndtere mediet, og fremvæksten af sociale medier og udviklingen af smarte mobiltelefoner skaber nye adfærdsmønstre.

For ordbogsudviklere indikerer undersøgelsens resultat, at det kan være svært at slå igennem med ordbøger på nettet alene i kraft af informationsværdi og almindelig efterspørgsel via søgemaskinerne.

5.3. Typiske forskelle mellem danske og svenske brugere

Der er konstateret forskellige grader af forskelle mellem brugen af ordbøgerne på den danske og svenske side, og det vil være rimeligt at antage, at det ikke skyldes tilfældigheder, men at det handler om konkrete forskelle.

Man kan bemærke, at mange tekniske ordbøger har næsten samme efterspørgsel i Danmark og Sverige, derimod er der flere ordbøger under ”Fagord” og ”Hverdagsord” med klare forskelle. De største forskelle foreligger imidlertid for ordbøgerne under temaet ”Slang”, hvor danskernes efterspørgsel ofte er større end svenskernes. Umiddelbart ser der således ud til at tegne sig et mønster på tværs af ordbøgerne.

Forskellene kan evt. afspejle danskeres og svenskeres forskellige behov, når de rejser eller flytter over på den anden side af sundet, f.eks. mht. arbejde eller ferie. Antallet af flyttede over sundet toppede desuden i 2008, samtidigt med at mange ordbøgers efterspørgsel også toppede.

Det er også fristende at karakterisere forskellene ud fra stereotyper om svenskere og danskere, at danskerne har hang til det underholdende og

kuriøse, mens svenskerne er mere seriøse og interesserer sig for de reelle sprogforskelle inden for f.eks. handelsvarer og falske venner.

Uden at gå ind i en sådan diskussion, kan man i hvert fald sige, at en svensk-dansk ordbog indholdsmæssigt bør adskille sig en del fra en dansk-svensk ordbog, hvis brugernes subjektive interesser skal sættes i fokus.

5.4. Udviklingen hos private og professionelle brugere

Resultaterne blotlægger lidt forskellige spørgeprofiler hos private og professionelle spørgere, men tendensen er den samme. En forskydning mod et mere fagligt emnevalg kan konstateres for begge bruger kategorier.

Den klart aftagende spørgefrekvens indikerer imidlertid faldende engagement og mindre vilje til at grave dybere. En bidragende årsag er naturligvis, at brugerne er blevet mere teknisk erfarne og søger andre steder hen, men det er næppe hele forklaringen, og det kan ikke udelukkes, at tendensen udtrykker et faktisk fald i brugerengagementet.

Hvad der især er vanskeligt at forstå er, at antallet af kvalificerede og rigtigt svære spørgsmål er faldet kraftigt i perioden (selv om de altså relativt set udgør en stigende andel af spørgsmålene som vist i dette studie). Dette er dog ikke et ukendt fænomen. En lignende udvikling er nemlig også set hos mange diskussionsfora for professionelle oversættere, f.eks. Norskjal. Man kan undre sig over, hvor oversætterne så går hen, når de behøver ekspert-hjælp, og hvordan det påvirker kvaliteten i deres oversættelsesarbejde, når de ikke søger kollegial hjælp i samme udstrækning som tidligere.

6. Konklusion

Foreliggende undersøgelse er interessant, fordi der ikke tidligere er publiceret statistik vedrørende ordbogsbrugeres efterspørgsel efter emnespecifikke ordbøger på nettet i et længere tidsperspektiv. Resultaterne viser, at ordbogsbrugernes præferencer over tid følger typiske mønstre, som kan adskille sig meget fra emne til emne, samt fra land til land. Kendetegnene for mange ordbøger er imidlertid tegn på en faldende tendens i brugeranvendelse og -engagement. I fremtiden kan der kræves flere virkemidler for at nå ud til brugerne.

Litteratur

- Kristensen, K. (2012): Tre svensk-danske ordbøger på nettet. I: *LexicoNordica* 19, 273-294.
- Lorentzen, H. & Theilgaard, L. (2012): Netordbøger – hvordan finder brugerne frem til dem, og hvad gør de, når de har fundet dem? I: *Nordiske studier i leksikografi* 11. Rapport fra Konferencen om leksikografi i Norden. Lund 24.-27.maj 2011, 418-430.
- Pálfi, Loránd-Levente (2011): *Leksikon over ordbøger og leksika*, 2. udgave. København: Frydenlund.

Poul Hansen
Statsautoriseret translatør, ph.d.
Box 4158
SE-227 22 Lund
poul@danska-svenska.se

Flerspråkig lexikografi och obeläktade språk – erfarenheter från finska Islex

Helga Hilmiðóttir & Nina Martola

The aim of the study is to explore some of the lexicographic problems editors encounter when they work on bilingual (or multilingual) dictionaries between unrelated languages such as Icelandic and Finnish. The problems that are used as examples are drawn from the Finnish version of the Icelandic-Scandinavian webdictionary Islex. In the article, we show how structural differences between the Scandinavian languages and Finnish call for special solutions in the Finnish edition. A special focus is on expression of ownership, problems that arise when lexical items are used without a context, and negations.

1. Inledning

Syftet med denna artikel är att dryfta lexikografiska problem som kan uppstå när man redigerar en ordbok mellan två obeläktade språk, närmare bestämt isländska och finska. Problemen blev aktuella när man våren 2012 inledde arbetet av den finska versionen av Islex som planeras vara färdig våren 2017.

Islex (www.islex.is) är en flerspråkig nätordbok mellan isländska och de övriga skandinaviska språken: danska, nynorska, bokmål, svenska och färöiska. Den isländska databasen är skapad av Árni Magnússon-institutet som också står bakom utformningen av ordboken. Varje enskilt land har sedan ansvarat för översättningen av uppslagsord och frasmaterial till respektive språk.

Ordboken har 50 000 uppslagsord och omkring 10 000 fasta och halvfasta fraser och 32 000 exempel (Sigurðardóttir m.fl. 2008). Enligt huvudredaktörerna var syftet bakom Islex att skapa en modern, aktuell ordbok mellan isländska och de andra skandinaviska språken (Sigurðardóttir m.fl. 2008). De existerande ordböckerna hade antingen få uppslagsord, var föråldrade eller hade inte den lexikografiska kvaliteten som användare kräver i dag.

Behovet av en finsk version är stort. Det finns en finsk-isländsk-finsk ordbok i tryckt form (Järvelä 2008), men den är betydligt mindre till omfånget än Islex och den har mycket korta artiklar. Inga exempel eller fasta fraser anges och boken saknar all grammatisk information.

2. Lexikografi mellan obesläktade språk

Ekvivalensproblematik är av naturliga skäl ett centralt ämne inom två- och flerspråkig lexikografi. Överlag verkar det dock vara semantiska frågor som dominerar i diskussionerna. Det är i och för sig inte oväntat eftersom ordböcker ju först och främst befattar sig med lexikala enheter. Den som arbetar med ordböcker mellan förhållandevis närbesläktade språk konfronteras inte heller som lexikograf med så avgörande strukturella skillnader mellan språken. Det är främst lexikografer från icke-europeiska språk som kommer in på problem föranledda av syntaktiska skillnader (t.ex. Kotze 1992, Moeimam & Steinhauer 2007, Prinsloo 2009).

I Hannesdóttir (2012) finns en åskådlig redogörelse för vilka slag av ekvivalensproblematik redaktionen för det svenska Islexprojektet varit tvungen att hantera. Den finska Islexredaktörerna har givetvis råkat ut för samma typer av ”traditionella” ekvivalensproblem. I sådana fall har de dock haft betydande hjälp av de andra målspråken och särskilt av svenska, eftersom de nordiska samhällena ändå är förhållandevis lika. Den verkligt svåra nöten att knäcka vad gäller finska som målspråk är emellertid de problem som uppstår på grund av de strukturella skillnaderna mellan isländska och finska, och vi ska i det följande ta upp några olika fall.

En viktig aspekt i ordbokssammanhang är målgruppen. Enligt presentationen på webbplatsen vänder sig Islex både till nordiska användare och till ”isländska användare som behöver information om ord och uttryck på de skandinaviska språken.” Hannesdóttir (2014) säger till och med att ordboken i första hand är avsedd som en produktionsordbok för islänningar, och avser förmodligen då i enlighet med NLO (1997) en tvåspråkig produktionsordbok avsedd för produktion på det främmande språket. I fråga om de andra skandinaviska språken har det förmodligen gått rätt bra att åstadkomma bidirektionalitet i Islex, men för finskans del måste man fråga sig om det inte är en orealistisk ambition.

Hannesdóttir (2014) slår fast att dikotomin mellan källspråk och målspråk först och främst gäller tryckta ordböcker och att *sekundärspråk* vore en lämpligare term än *målspråk* när det gäller elektroniska ordböcker. I en ordbok som är sökbar också med avseende på sekundärspråket ställs det betydligt högre krav på detta i fråga om täckningsgrad och relevans. Det medför nya metodologiska utmaningar inom lexikografin. Att göra finska till ett fungerande sekundärspråk i en nordisk ordbok är förenat med mycket stora svårigheter, i all synnerhet om man tänker sig att också användare med enbart nybörjarkunskaper i finska ska kunna tillgodogöra sig informationen. (För en inblick i tvåspråkigt ordboksarbete med finska som källspråk hänvisas till Seppinen (1992).

3. Analys av några syntaktiska problem

I detta kapitel diskuterar vi tre problemområden som redaktionen har konfronterats med under arbetets gång: ägande, asymmetrisk utbytbarhet och optionellt nekande satser. De är representativa för den typ av problematik det är frågan om, och både ägarkonstruktioner och optionellt nekande satser är dessutom frekventa i ordboken.

3.1. Ägande

Ägande i isländska uttrycks med tre olika verb: *að eiga* 'äga, ha', *að hafa* 'ha' och *að vera með* 'ha till sitt förfogande, ha med sig'. Samtliga verb ingår i en vanlig SVO-ordföljd där objektet står i ackusativ: *hann á hund* 'han har en hund', *hún hefur tíma* 'hon har tid' och *konan er með gleraugu* 'kvinnan har glasögon'. Verbet *eiga* 'äga' används dels för att uttrycka ägandet av konkreta saker som hus, pennor och katter, och dels för att uttrycka släkt- och kärleksrelationer (*að eiga bróður* 'ha en bror', *að eiga konu* 'ha en fru'). Användningen av *að hafa* är däremot rätt begränsad och används mest när det gäller immateriella företeelser som tid och intresse (*hafa tíma* 'ha tid', *hafa áhuga* 'ha intresse'). (ÍO 2004.)

På finska finns det ett verb med betydelsen 'äga', *omistaa*. Det används enligt den finska ordboken *Kielitoimiston sanakirja* för att ange att egendom (särskilt värdefull sådan) tillhör någon. För att mer allmänt uttrycka att nå-

gon har något används i finska en typ av existentialkonstruktion. Verbet är *olla* 'vara' och ägaren anges med ett adverb med nomenet i adessiv. Det ägda fungerar som grammatiskt subjekt (om än med vissa atypiska drag): *hänellä on koira* (ordagrant: 'hos honom är en hund') (VISK § 895, se även t.ex. Häkkinen 2005).

ég á hund
jag.NOM ha.1.SING hund.ACK
'jag har en hund'

minulla on koira
jag.ADESS vara.3.SING hund.NOM
'jag har en hund'

Att ange finsk ekvivalent i artikeln *eiga* (Bild 1) är alltså mer problematiskt än för de andra språken. Verbet *omistaa* är det enda verbet som kan anges utan kontext men det motsvarar sällan *eiga*. Således fungerar verbet när ägandet gäller t.ex. fastigheter, landområden, aktier o.likn. (*omistaa talo* 'äga ett hus') men inte när det gäller att ha pengar, familj eller småsaker som telefoner och klockor. I sådana fall används verbet *olla* 'vara', men det kan däremot inte anges som ekvivalent utan att visas i kontext eftersom det kontextlöst har betydelsen 'vara'.

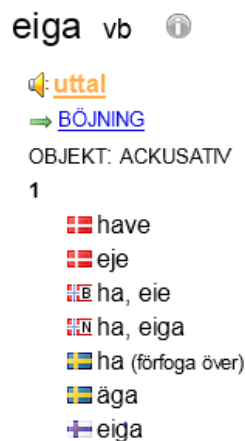


Bild 1: Uppslagsordet *eiga* med ekvivalenter.

En eventuell lösning är att skriva ut en grammatisk förklaring och markera den som sådan (jfr. Sigurðardóttir 2008). En sådan förklaring skulle dock ha en begränsad nytta för en isländsk användare av Islex som vill lära sig finska. Möjligen kunde upplysningen kompletteras med en länk till ett fönster med en utförligare förklaring och exempel både på isländska och finska. Tekniskt är det här fullt möjligt men eftersom vi ännu inte har börjat diskutera de verkligt svåra fallen för finskan med de isländska huvudredaktörerna är detta ett problem som återstår att lösa. Ytterligare ett alternativ vore att lägga in adessivformen i ekvivalenten: *olla [jollakulla]*. Också en ekvivalent kan förstås länkas till grammatiska förklaringar.

Det som ytterligare skapar problem i exempel som anger ägande är att en stor del av verbexemplen i Islex består av subjektlösa sats. Det fungerar lexicografiskt för de andra språken i Islex, eftersom ägandet ingår som en betydelsekomponent i verbet/verbfrasen. Följande exempel anges för verbfrasen *vera með* 'ha på sig':

vera + með

***vera með* <eldspýtur>**







-  have <tændstikker>
-  være i besiddelse af <tændstikker>
-  ha <fyrstikker>
-  ha <fyrstikker>
-  ha <tändstickor> (på sig eller med sig)
-  hava <sváulpinnar> uppi á sær

Bild 2: Uppslagsordet *vera með* med ekvivalenter.

Som Bild 2 visar anges verbfrasen *vera með* med en subjektlös sats *vera með* <eldspýtur> 'ha <tändstickor>'. I finska måste ägarrollen anges explicit och vårt tentativa förslag är att man i översättningen av *vera með* <eldspýtur> tillägger ett pronomen i tredje person *hänellä* 'hos henne/honom' i hakparenteser. Runda parenteser fungerar inte eftersom de också används för att visa alternativa ord som kan uteslutas som t.ex. negationer (jfr avsnitt 3.3):

vera með <eldspýtur>
 vara.INF med tändstickor.PL.ACK
 'ha tändstickor'

<hänellä> on <tulitikkuja>
 hen.ADESS vara.3.SING tändstickor.PL.PART
 'hen har tändstickor'

Ytterligare en möjlighet är att markera det tillagda ordet i en annan färg. Oavsett vilken markering man använder innebär lösningen dock en asymmetri, eftersom källspråkets infinitiva sats *vera med 'ha'* motsvaras av en finit sats där finskans verbform är *on 'är'*, inte *olla 'vara'*.

3.2. Existentialkonstruktion och utbytbara element

Det finska verbet *olla 'vara'* används i många fall där isländska och de andra nordiska språken har ett betydelseäldre verb eller en verbförbindelse för att ange en omständighet eller ett skeende. Ett exempel på detta är *fara fram 'äga rum'* som i finska ofta motsvaras av en existentialkonstruktion (Bild 3).







<útskriftin> fer fram <í stóra salnum>
 <afslutningshøjtideligheden> finder sted <i den store sal>
 <eksamensfesten> findet sted <i den store salen>
 <eksamensfesten> går føre seg <i den store salen>
 <examenshögtiden> äger rum <i aulan>
 <próvhandanin> fer fram <í stóra salinum>
 <valmistujaisjuhla on isossa salissa>

Bild 3: Frasen *fara fram* med översättningar till samtliga språk¹.

I detta och liknande fall blir de utbytbara elementen inom vinkelparenteser problematiska. För att verbet *olla* ska få innebörden 'äga rum' måste två villkor vara uppfyllda; det måste finnas en platsangivelse och subjektet måste vara ett nomen som anger skeende. I satsen *projektorin on isossa salissa* har *olla* betydelsen 'vara, finnas' ('projektorn är/finns i aulan'). Både

¹ Redaktionen har sedermera stannat för verbet *pitää 'hålla'* i stället för *olla 'vara'*.

valmistujaisjuhla och *isossa salissa* är naturligtvis på samma sätt utbytbara också i finska, men läggs elementen inom vinkelparenteser kommer det att skapa intrycket att *on* ensamt betyder 'äger rum'. Om finska hade varit med från början hade vi som redigerar den finska delen lagt fram ett förslag om att användningen av vinkelparenteser skulle begränsas betydligt och i detta fall strykas, men eftersom ordboken redan har publiceras på de övriga språken föreslår vi inte några åtgärder här.

3.3. Nekande konstruktioner

En annan konstruktion som vållar problem för den finska Islexredaktionen är nekande konstruktion. På de nordiska språken görs satser nekande med hjälp av ett nekningsord, *han såg katten* respektive *han såg **inte** katten*. Därmed är det i regel oproblematiskt att ange alternativt nekande sats, och det förekommer rätt frekvent i Islex, som i Bild 4.

halda + út

halda <þetta> (ekki) út

OBJEKT: ACKUSATIV

🇳🇵 (ikke) kunne holde <det her> ud;

🇳🇵 (ikke) orke <det her>

🇩🇪 (ikke) holde ut <dette>

🇩🇪 (ikke) orke <dette>

🇳🇵 (ikkje) halda ut <dette>

🇳🇵 (ikkje) orka <dette>

🇸🇪 (inte) stå ut med <det här>;

🇸🇪 (inte) orka med <det här>

🇫🇮 eg haldi <hetta> ikki út

Bild 4: Nekande sats som alternativ.

I finska görs en sats nekande med hjälp av en verbkonstruktion. Dessutom påverkas objektets kasus vid verb som inte har partitivobjekt också i jakande sats.

hän näki kissan
 hen se.PRET.3.SING katt.ACK.
 'hen såg en katt/katten'

hän	ei	nähnyt	kissaa
hen	NEK.HJV.3.SING	se.PCP.SG	katt.PART
'hen såg inte katten'			
'hen såg ingen katt'			

Neknings verbet böjs i person och participet varierar i standardfinska mellan singularis och pluralis (*en nähnyt* 'jag såg inte', *emme nähneet* 'vi såg inte', *he eivät nähneet* 'de såg inte'). Exemplet ovan anger förfluten tid. I presens används en särskild (oböjlig) nekningensform av huvud verbet i stället för particip (*en näe*, *emme näe*, *he eivät näe* 'jag/vi/de ser inte'). (VISK § 108).

Ytterligare en komplikation är att finska saknar nekande infinitiv. Jakande infinitiv kan uttryckas på motsvarande sätt som i skandinaviska språk *nähdä kissa* 'se en katt'. Däremot låter sig inte den nekande motsvarigheten *inte se en katt* utan vidare översättas. Det finns en verbförbindelse bildad med hjälp av s.k. MA-infinitiv, <olla -mAttA> som utgör en teoretisk motsvarighet (t.ex. *olla näkemättä kissaa*), men den fungerar inte lexikografiskt, eftersom satserna ofta får en klang av avsiktlighet. Det går alltså under inga omständigheter att på ett enkelt sätt ge en finsk översättning för *halda* <*petta*> (*ekki út*, som för de övriga språken (bild 4 ovan). På ordplanet finns det i detta fall ett utmärkt verb *kestää* 'stå ut', och lösningen den finska redaktionen stannat för är att ge frasen i finit form och att ge två separata översättningar för att få variationen mellan jakande och nekande sats tydlig:

[hän] *kestää* <tämän>
 [hän] *ei kestä* <tätä>

Om den isländska frasen inte hade innehållit det optionella (*ekki*) kunde infinitiv till nöds ha varit möjligt också på finska.

kestää <tämä>

Som exemplen ovan visar, föreslår vi att den finska versionen av Islex anger två olika motsvarigheter för optionellt nekande satser. Den förra i jakande betydelse och den senare med negation.

4. Avslutande diskussion

I det föregående har vi diskuterat några problem som uppkommit i arbetet med finska Islex och som härrör från att finska skiljer sig strukturellt från isländska och från de andra målspråken. I en del fall har redaktionen fått stanna för en lösning som inte känts helt tillfredsställande.

Till en del handlar det om metodologiska frågor. Hade finska varit med från början kunde hänsyn ha tagits – åtminstone i viss utsträckning – till dess särskilda behov. Artikelstrukturen kunde t.ex. ha utformats lite annorlunda och exemplen kunde i högre grad ha getts i finit form. Nu kom finska in när danska, svenska och delvis norska redan förelåg som målspråk. Varje ändring i källspråksfraserna skulle därmed kräva åtgärder i de andra målspråken, och de andra redaktionerna fungerar inte längre aktivt.

Att använda en gemensam källspråksbas för ett stort antal målspråk är en rätt vanlig metod i lexikografiska sammanhang. Så gjordes t.ex. svenska Lexin och så görs för närvarande också de finska invandrarspråksordböckerna. (Också i det svenska Lexinprojektet hade för övrigt den finska redaktionen problem [Paula Ehrnebo, muntlig kommunikation]). En gemensam källspråksbas för flera målspråk gör naturligtvis det lexikografiska redigeringsarbetet enklare på många vis. De finska redaktörerna har som ovan nämdes haft stor hjälp av de andra målspråken och särskilt av svenska, och arbetet har i själva verket gått snabbare än redigeringen av de centralskandinaviska målspråken. Finska och svenska står varandra nära kulturellt och samhälleligt.

Så länge en ordbok är monodirektionell fungerar metoden med en gemensam bas hyfsat. Om ordbokens målspråk² är användarens modersmål kan man tillåta sig att i viss mån ge avkall på det idiomatiska, men frågan är förstås i hur hög utsträckning. Kunde man t.ex. i det ovan angivna tändsticksfallet helt enkelt ge *olla* <*tulitikkuja*> och i *halda út*-fallet (*ei*) *kestäi* <*tätä*> och inte vara så petig? Finska Islexanvändare vet att tändstickor är något man har och inte något man är och de vet hur finska nekande satser

² Här används *målspråk* i traditionell lexikografisk bemärkelse. När det gäller t.ex. invandrarordböcker som Lexin är det ju i språkinläringbemärkelse ordbokens källspråk som är användarnas målspråk, det språk de ska lära sig. Se också Hannesdóttir (2014) för termerna *målspråk* och *sekundärspråk*.

konstrueras. Den oortodoxa översättningen skulle hjälpa dem att tolka och/eller lära sig den isländska frasen och det kunde ju anses räcka. Har man däremot ambitionen att finska ska vara ett reglerätt sekundärspråk (jfr Hannesdóttir 2014) borde man ställa stora krav på idiomacitet.

Å andra sidan har vi den aspekt som Hannesdóttir (2012, 2014) tar upp, att elektroniska två- eller flerspråkiga ordböcker i regel är (och i vart fall förväntas vara) sökbara både för käll- och målspråk, och för att en finsk användare ska kunna utföra sökningar måste det finska målspråket också av den orsaken vara idiomatiskt. Men det gäller förstås i första hand frassökning. Sökning på ord berörs ju inte av strukturella skillnader i någon nämnvärd grad.

I förordet till Islex anges det att ordboken riktar sig till fler grupper av användare, främst då talare av målspråken, men ”å andra sidan isländska användare som behöver information om ord och uttryck på de skandinaviska språken.” Också Hannesdóttir (2014) anger islänningar som en målgrupp. Riktningen kan tänkas fungera för de andra målspråken men definitivt inte för finska. För att isländska användare ska kunna tillgodogöra sig information om finska i den isländsk-finska ordboksdelen borde den se annorlunda ut.

Litteratur

Ordböcker

ÍO = *Íslensk orðabók* (2004). Reykjavík: Edda.

Islex. <www.islex.is> (september 2015).

Järvälä, Tuomas (2008): *Suomi-islanti-suomi sanakirja*. Helsinki: Gaudeamus.

Kielitoimiston sanakirja (2014). Kotimaisten kielten keskuksen verkkojulkaisuja 35. Helsinki: Kotimaisten kielten keskus. <www.kielitoimistonsanakirja.fi/netmot.exe?motportal=80> (september 2015).

Lexin – lexikon för invandrare. Institutet för språk och folkminnen och Kungliga tekniska högskolan. <lexin.nada.kth.se/lexin> (september 2015).

NLO (1997) = Henning Bergenholtz, Ilse Cantell, Ruth Vatvedt Fjeld, Dag Gundersen, Jón Hilmar Jónsson & Bo Svensén: *Nordisk leksikografisk ordbok*. Oslo: Universitetsforlaget.

Norstedts stora engelska ordbok (2011). Stockholm: Norstedts.

Nordiske Studier i Leksikografi 13 (2016): 143-153

Sekundärlitteratur

- Hannesdóttir, Anna Helga (2012): Ekvivalensrelationer i tvåspråkig lexikografi. I: *LexicoNordica* 19, 39–58.
- Hannesdóttir, Anna Helga (2014): Lemman och ekvivalenter i nya roller – en reviderad ordbokstypologi. I: *Nordiske studier i leksikografi. Rapport fra Konferanse om leksikografi i Norden*. Oslo 13.–16. august 2013, 193–211.
- Häkkinen, Kaisa (2005). Finska. I: *Norden språk med rötter och fötter*. København: Nordisk ministerråd. <eplads.norden.org/nordenssprak/kap2/2b/05.asp> (september 2015).
- Kotze (1992): The Lexicographical Handling of Grammatical Equivalence. I: *Lexikos* 2. <reference.sabinet.co.za/webx/access/journal_archive/16844904/406.pdf> (september 2015).
- Moeimam, Susi & Hein Steinhauer (2007): The Dutch-Indonesian Dictionary Project. I: *International Journal of Lexicography*, Volume 20, 275–293.
- Prinsloo, D.J. (2009): Current Lexicography Practice in Bantu with Specific Reference to the Oxford Northern Sotho School Dictionary. I: *International Journal of Lexicography* 22, 151–178.
- Seppinen, Hannele (1992): Finska för utlänningar – uppgifter om ordböjning i den finsk-norska ordboken. I: *Nordiske Studier i Leksikografi* 1, *Rapport fra Konferanse om leksikografi i Norden* 28.-31. mai 1991. 436–440.
- Sigurðardóttir, Aldís m.fl. (2008). Islex - An Icelandic-Scandinavian Multilingual Online Dictionary. I: *Proceedings of the XIII Euralex International Congress: Barcelona, 15-19 July 2008, Sèrie activitats* (20), 529–539.
- VISK = Hakulinen, Auli, Maria Vilkuna, Riitta Korhonen, Vesa Koivisto, Tarja Riitta Heinonen & Irja Alho (2004): *Iso suomen kielioppi*. Helsinki: Suomalaisen Kirjallisuuden Seura. <scripta.kotus.fi/visk URN:ISBN:978-952-5446-35-7> (september 2015).

Helga Hilmisdóttir
 universitetslektor, fil.dr.
 Helsingfors universitet
 PB 14 (Unionsgatan 40)
 FI-00014 Helsingfors universitet
 helga.hilmisdottir@helsinki.fi

Nina Martola
 avdelningsföreståndare, fil.dr.
 Institutet för de inhemska språken
 Hagnäskajen 6
 FI-00100 Helsingfors
 nina.martola@sprakinstitutet.fi

Bruk av digitale ordbøker i norsk skule

Marit Hovdenak & Ingunn Indrebø Ims

This article presents the material from an online survey in 2014 which reveals if, how and in what way teachers and their students in the Norwegian schools use and relate to monolingual online dictionaries in Norwegian. 487 teachers and 5172 students from all over Norway responded to the survey. The findings are interesting with respect to several of the objectives of The Language Council of Norway, among others the responsibility to promote the two Norwegian written standards Bokmål and Nynorsk in the classrooms and ensure that the dictionaries are up to date.

1. Innleiing

Ordbøker er kjelder til kunnskap om norsk språk. Eit viktig formål for mange ordbøker og ordlister i Noreg er at dei skal gje informasjon om offisiell rettskriving for bokmål eller nynorsk til elevar og lærarar. I skulen er ordbøker ein type læremiddel, men det er lite faktakunnskap om korleis elevar bruker ordbøker i ulike situasjonar, og kva for ordbøker som er i bruk i skulen. I 2008 gjennomførte leksikografane Knut Karlsen og Dagfinn Rødningen ei lærarundersøking i vidaregåande skule for å sjå nærare på i kva grad einspråklege ordbøker blir nytta i norskfaget (Karlsen & Rødningen 2008). Dei skriv (sst.:93): "Ordlister blir rekna for viktige hjelpemiddel i skolesamanheng, også i morsmålet, men systematisk innhenta kunnskap om kva slags bøker som blir nytta, korleis ein nyttar dei og kvifor, har vi ikkje mykje av." Dei legg òg til at det er viktig med betre innsyn i korleis skulen nyttar ordlister og ordbøker (sst.). Undersøkinga til Karlsen og Rødningen er ei lærarundersøking i vidaregåande skule, og i spørsmåla blir papirbaserte og nettbaserte ordbøker sette under eitt. Frå dei siste åra finst det eit par undersøkingar om bruken av fleispråklege og einspråklege digitale ordbøker i engelskundervisninga (Graedler 2015) og bruken av ordbokressursar for minoritetsspråklege (Rambøll 2012 og Bjørneset 2015), men det finst ingen studiar som ser nærare på kva for norskspråklege digitale ordbøker som er i

bruk i skulen både frå eit lærar- og elevperspektiv. Difor fekk Språkrådet hausten 2014 gjort ei landsomfattande undersøking om bruk av digitale ordbøker i norsk skule.¹ Undersøkinga tek primært for seg bruken av digitale ordbøker i skulesamanheng, men ho gjev òg generelle svar på i kva grad elevar i skulen nyttar papirordbøker, og om bruken av ulike typar ordbokressursar endrar seg gjennom skuleløpet. Undersøkinga omfattar elevar og lærarar på mellom- og ungdomstrinnet og i den vidaregåande skulen. 487 lærarar og 5172 elevar frå heile landet svarte på spørjeskjema. Denne artikkelen presenterer dei viktigaste resultatane og drøftar korleis undersøkinga er nyttig for Språkrådet i vidare arbeid.

2. Formålet med undersøkinga

Mange kommunar og fylkeskommunar har som skuleeigarar kjøpt inn digitale ordbokressursar til skulane, men vi veit ikkje om eller når elevane bruker dei, eller om dei heller slår opp i andre ordbøker dei finn på nettet. Formålet med undersøkinga var difor å finne ut kva for digitale ordbøker² (bokmål og nynorsk) som er i bruk i skulen, og i kva grad dei blir nytta i ulike situasjonar. Både elevar og lærarar fekk spørsmål om dei bruker ordbokressursar heime, i undervisninga, i rettleiararbeid og under heildagsprøver som eksamen og tentamen. Dei fekk òg spørsmål om kva for digitale ordbøker som er i bruk i dei ulike situasjonane.

3. Språkrådet og ordbøker

Språkrådet er opptekte av korleis språket fungerer på ulike samfunnsområde og ser ordbøker som ein viktig del av den samla språkpolitikken. Når det gjeld skulen, arbeider Språkrådet med ordbøker på særleg to felt: Språkrådet står bak standardordbøkene *Bokmålsordboka* (BOB) og *Nynorskordboka* (NOB) saman med Universitetet i Oslo, og Språkrådet er ansvarleg for ei

¹ Undersøkinga vart gjennomført av TNS Gallup på vegner av Språkrådet. Ingunn Indrebø Ims og Kristin Solbjør frå Språkrådet stod for arbeidet med spørjeskjema.

² I undersøkinga og i denne artikkelen omfattar termen *digitale ordbøker* både lisensierte digitale ordbokressursar og ordbøker som er gratis tilgjengelege på Internett.

lovfesta godkjenningsordning for ordlister og ordbøker til skulebruk, ei ordning som skal reviderast for å svare til dagens krav.

3.1. BOB og NOB

BOB og NOB er digitale definisjons- og rettskrivingsordbøker, som òg finst i papirutgåver. Nettordbøkene er gratis tilgjengelege for alle, og dei har hatt ein aukande bruk, med tredobling i nettsøka frå 2010 til 2014. Då vart det gjort 49 millionar søk i BOB og 34 millionar søk i NOB (Språkrådet 2014:4). BOB og NOB er utarbeidde i leksikografamiljøet ved Universitetet i Oslo i samarbeid med Språkrådet og kom ut første gongen i 1986. Dei har komme i fleire utgåver med oppdateringar i rettskriving og med nye ord, og ordbøkene er no knytte til fullformslister i rettskrivingsbasane Norsk ordbank, som er tilgjengelege på nettsidene til Universitetet i Oslo og også frå Norsk språkbank. Ordbøkene er den viktigaste kanalen Språkrådet har for å formidle informasjon om gjeldande offisiell rettskriving til publikum.

3.2. Godkjenningsordninga

Rettskrivinga i bokmål og nynorsk har vore eit sentralt språkpolitisk emne i Noreg, og på 1900-talet var det fleire store rettskrivingsreformer. Frå 1950-åra var det Norsk språknemnd og etterfølgjaren Norsk språkråd som stod for større og mindre endringar i rettskrivinga. Oppdaterte ordlister var viktig, særleg for å nå fram til skulane på ein effektiv måte. Vi har ei offentleg godkjenningsordning for skuleordlister og -ordbøker, ei ordning som tidlegare inngjekk i ei generell godkjenningsordning for lærebøker, oppheva i år 2000. Godkjenningsordninga har no heimel i den sentrale opplæringslova under punktet "Lærebøker og andre læremiddel": "Ordlister og ordbøker til skulebruk skal godkjennast av Språkrådet" (opplæringslova § 9-4, fjerde ledd). Godkjenninga inneber ein nitid kontroll for å sikre at ordlista eller ordboka er i samsvar med gjeldande rettskriving. I samarbeid med bl.a. forlaga har Språkrådet freista å skissere endringar i ordninga. Språkrådet vil halde fast på at ordlistene og ordbøkene skal opplyse om alle valfrie former og vil leggje til rette for ein automatisk kontroll mot Norsk ordbank. Mindre oppdateringar må kunne gjerast enkelt.

4. Undersøkinga ”Digitale ordbøker i bruk”

4.1. Metode og utval

Undersøkinga ”Digitale ordbøker i bruk” vart gjennomført hausten 2014 av meiningsmålingsfirmaet TNS Gallup. Undersøkinga vart gjennomført blant elevar og lærarar på mellomtrinnet (5.–7. trinn), ungdomstrinnet (8.–10. trinn) og vidaregåande skule. 1000 skular med mellomtrinn og 1000 skular med ungdomstrinn vart inviterte til å delta. I tillegg vart alle vidaregåande skular i Noreg inkluderte i utvalet (429) (TNS Gallup og Språkrådet 2014).³ Svarprosenten i grunnskulen var 11 %, i vidaregåande skule 25 %. Til saman 5172 elevar og 487 lærarar svarte på undersøkinga. Låg svarprosent, særleg i grunnskulen, gjer at resultatata ikkje kan seie noko om *utbreiing* av digitale ordbøker, men dei seier noko om korleis elevar *bruker* digitale ordbøker i dei skulane der desse ressursane er tekne i bruk.

Elevar og lærarar fekk ulike spørjeskjema. Elevskjemaet inneheldt 17 spørsmål om: 1) bruk av datamaskin i undervisninga, 2) kjennskap til digitale ordbøker, 3) bruk av digitale ordbøker generelt. I tillegg fekk elevane i vidaregåande skule to spørsmål om bruk av digitale ordbøker under heildagsprøver. Lærarskjemaet inneheldt 15 spørsmål om: 1) informasjon om ordbøker i undervisninga, 2) tilgang til digitale ordbøker på skulen, 3) bruk av digitale ordbøker i undervisning og eige arbeid, 4) ordbøker under heildagsprøver. Fordi spørsmåla både til lærarar og elevar i hovudsak dreidde seg om kva for digitale ordbøker dei kjende til og nytta i ulike situasjonar, var svaralternativa på alle spørsmåla gjevne på førehand ved fleirval. På dei fleste spørsmåla kunne elevane og lærarane likevel krysse av på alternativet ”anna” og fyller inn sjølve, dersom dei hadde andre svar.

Den vidare gjennomgangen av resultatata frå undersøkinga er strukturert etter tema i dei ulike spørjeskjemaa.

³ For meir informasjon om metode, utval, veking og feilmargar, sjå rapporten *Digitale ordbøker i bruk. Undersøkelse blant elever og lærere på mellom- og ungdomstrinnet og i den vidaregåande skolen* (2014).

4.2. Elevane

4.2.1. Bruk av datamaskin i undervisninga

Fordi undersøkinga primært tek for seg bruk av digitale ordbøker, var det interessant å få vite noko om elevane sin tilgang til datamaskin på skulen og i undervisninga. Utreidd bruk av nettbaserte ordbokressursar krev naturleg nok tilgang til utstyr som gjer det mogleg å nyttiggjere seg slike ressursar. Tilgangen til eiga datamaskin på skulen vil altså ha samanheng med om elevane hovudsakleg nyttar papirbaserte eller nettbaserte ordbøker. Innleiingsvis i spørjeskjemaet får elevane difor spørsmål om dei har tilgang til eiga datamaskin på skulen som berre dei bruker. Resultata frå undersøkinga viser at bruken av eiga datamaskin i undervisninga aukar markant med alderen til elevane. Medan 10 % av elevane på mellomtrinnet bruker datamaskin ”alltid” eller ”ofte” i norskundervisninga, er det tilsvarande talet for elevar i vidaregåande skule 70 %. Omtrent alle elevar i vidaregåande skule svarer òg at dei har tilgang til eiga datamaskin på skulen som berre dei bruker. Dette gjeld nesten ingen av elevane på mellomtrinnet. Dette er ikkje overraskande tal, men bruken av datamaskin i undervisninga heng saman med tilgang til digitale læringsressursar, og dette viser også at tilgangen til digitale ordbokressursar versus papirordbøker endrar seg gjennom utdanningsløpet. Elevar i vidaregåande vil i større grad bruke digitale ordbøker enn det elevar på lågare trinn gjer. Dei fleste papirordbøker som blir nytta i skulen, er godkjende av Språkrådet til skulebruk, medan fleire av dei digitale ordbokressursane som blir nytta, enda ikkje er det. Ein slik overgang frå papirordbøker til digitale ordbøker gjennom utdanningsløpet kan difor føre til at enkelte elevar har meir oppdatert informasjon om rettskrivingsnormene tidleg i utdanningsløpet (gjennom dei godkjende papirordbøkene) enn dei har seinare.

4.2.2. Kjennskap til digitale ordbøker

Det er fleire elevar på vidaregåande skule (50 %) enn på mellom- (28 %) og ungdomstrinnet (38 %) som svarer at dei har fått kjennskap til digitale ordbøker frå læraren. Men elevar på mellomtrinnet rapporterer i større grad enn dei andre gruppene at dei har fått kjennskap til digitale ordbøker frå

foreldra sine. Gjeldande læreplan i norsk har som eitt av kompetansemåla etter 7. trinn at elevane skal kunne ”bruke forskjellige former for digitale og papirbaserte ordbøker”. Når berre 28 % av elevane på mellomtrinnet meiner at dei har lært noko om digitale ordbøker på skulen, samstundes som dei har fått informasjon frå foreldra om dette, kan det tyde på at lærarane ikkje følgjer opp kompetansemåla i læreplanen slik dei skal. Men det kan òg tyde på at elevane treng slik opplæring tidlegare enn læreplanen legg opp til.

Når det gjeld spørsmålet om kva for digitale ordbøker elevane kjenner til, viser resultatane at BOB (31 %) og NOB (28 %) og ordbøker frå Ordnett (29 %) og iFinger (25 %) er dei mest kjende ordbøkene. Desse tala gjeld heile elevgruppa sett under eitt. BOB og NOB er som tidlegare nemnt utvikla av Språkrådet i samarbeid med Universitetet i Oslo og er tilgjengelege gratis på Internett. Ordbøker frå Ordnett og iFinger er digitale ordbokressursar som gjerne er kjøpte inn av fylkeskommunen, og som blir nytta av skulane i enkelte fylke.

4.2.3. Bruk av digitale ordbøker

To av tre elevar i vidaregåande skule svarer at dei nyttar dei digitale ordbøkene som læraren har anbefalt. På spørsmål om kva for digitale ordbøker skulen eller læraren har anbefalt, fordeler svara seg slik: Ordbøker frå Ordnett (33 %), NOB (31 %), iFinger (31 %), BOB (31 %). På mellom- og ungdomstrinnet svarer fleirtalet at dei *ikkje* nyttar digitale ordbøker som er anbefalte av læraren. Likevel seier elevar på mellom- (36 %) og ungdomstrinnet (49 %) at dei sjølve har søkt etter slike ressursar på Internett. 5 % av elevane på mellomtrinnet svarer at dei nyttar digitale ordbøker kvar veke, medan 33 % av elevane på vidaregåande skule svarer det same. Dersom ein ser på dei som svarer ”kvar veke” og ”kvar månad” under eitt, er forskjellen 13 og 52 %. Ein annan tendens er at elevar med eit anna morsmål eller førstespråk enn norsk (34 %) nyttar digitale ordbøker hyppigare enn elevar med norsk som førstespråk (21 %).

Sjølv om svært mange elevar i vidaregåande skule svarer at dei nyttar digitale ordbøker ”kvar veke” eller ”kvar månad”, er det likevel slik at fleire

nyttar papirordbøker (58 %) enn digitale ordbøker (44 %) til eksamen. 11 % svarte at dei ikkje nyttar ordbok på eksamen.⁴

4.3. Lærarane

4.3.1. Informasjon om ordbøker i undervisninga

Fordi det er eit kompetansemål i læreplanen i norsk at elevane skal ha kjennskap til og kunne nytte digitale ordbøker når dei er ferdige med mellomtrinnet, fekk lærarane spørsmål om når dei meiner det er relevant å introdusere elevane for ordbøker i undervisninga. Her viser resultatane at lærarane er på linje med målsetjinga i læreplanen; dei fleste lærarane svarer at dette bør skje i undervisninga på 4. og 5. trinn.

4.3.2. Tilgang til digitale ordbøker på skulen

Nær ni av ti lærarar i vidaregåande skule svarer at deira skule har tilgang til norskspråklege digitale ordbøker. På mellomtrinnet svarer om lag halvparten av lærarane at dei ikkje veit om skulen har tilgang til slike ordbokressursar. Trass i at det er lærarane på mellomtrinnet som etter kompetansemåla skal introdusere elevane for desse ressursane, tyder svara på at lærarane i denne gruppa er mindre medvitne om dette enn lærarar på ungdomstrinnet og i vidaregåande skule.

4.3.3. Bruk av digitale ordbøker i undervisning og eige arbeid

Av dei lærarane som svarer at skulen har tilgang på norskspråklege digitale ordbøker, svarer eit klart fleirtal at dei nyttar desse i undervisninga. Dette gjeld i større grad lærarar på ungdomstrinnet og vidaregåande skule enn på mellomtrinnet. Det kan òg sjå ut til at lærarar ved skular som har nynorsk som opplæringsmål eller hovudmålform, i større grad nyttar digitale ordbøker i undervisninga enn lærarar på skular der bokmål er mest nytta.

⁴ På dette spørsmålet var det mogleg å gje fleire svar.

Heile 82 % av lærarane i vidaregåande skule svarer at dei bruker digitale ordbøker når dei rettar elevarbeid. Til samanlikning gjeld dette 67 % av lærarane på ungdomstrinnet og 38 % på mellomtrinnet.

Som tidlegare nemnt skal ordlister og ordbøker til skulebruk godkjennast av Språkrådet, noko som skal sikre at ordlista eller ordboka er oppdatert og i tråd med gjeldande rettskriving for nynorsk eller bokmål. Denne ordninga omfattar både digitale ordbøker og papirordbøker. I undersøkinga svarer heile 27 % av lærarane at dei ikkje veit om dei ordbøkene skulen har tilgang til, er godkjende og i tråd med offisiell norsk rettskriving. Når vi veit at det finst digitale ordbøker på marknaden som ikkje er godkjende etter gjeldande rettskriving, og som vert nytta i skulen, er usikkerheita i svara frå lærarane eit signal om at paragrafen i opplæringslova må bli tettare følgd opp. Både elevar og lærarar bør vere trygge på at dei læremidla dei tek i bruk, er oppdaterte og i tråd med gjeldande rettskrivingsnorm.

5. Konklusjonar av undersøkinga – korleis kan Språkrådet bruke resultata?

Dette er ei gransking i stor skala av bruken av ordbøker blant elevar og lærarar, om kva for ressursar som er utbreidde, og om bruk i ulike situasjonar. Undersøkinga viser at det er eit ganske stort tilbod av ordbokressursar. Gjennom undersøkinga har vi dessutan fått ny kunnskap om i kva grad Språkrådet sine ordbøker, BOB og NOB, er i bruk i ulike skulesituasjonar. At ordbøkene er mykje brukte i skulen, av både elevar og lærarar, viser at arbeidet med å vedlikehalde og utvikle desse ordbokressursane er nødvendig. Dei mest brukte ordbøkene og ordlistene elles er godkjende av Språkrådet, dei fleste som papirutgåver, mens fleire òg inngår i nettportalar. Dessutan er det enkelte ordbøker i bruk som ikkje er godkjende av Språkrådet. Det har hendt at Språkrådet har fått klage på at det er store feil i ikkje-godkjende elektroniske ordlister. Denne situasjonen minner oss om at justeringar av godkjeningsordninga må komme på plass.

Resultatet som viser at BOB og NOB er mindre brukte ved eksamen enn elles, har medverka til at Språkrådet har gått ut med informasjon til skulane om at desse ordbøkene òg kan brukast ved eksamen ved ei spesiell ordning

som ligg innanfor regelverket. Ved eksamen og tentamen er regelen vanlegvis at elevane ikkje skal ha tilgang til Internett.

Språksamlingar og ordbøker er viktige ressursar for god språkrøkt og formidling av dei offisielle skriftnormene. Undersøkinga har vist at standardordbøkene BOB og NOB har ei sterk stilling i skuleverket. Det er viktig at dei blir haldne à jour, og no er det òg eit stort behov for revisjon. For den daglege drifta og vedlikehaldet av dei er det avgjerande å ha tilgang til fagkompetanse i nordisk språkvitskap, leksikografi og IKT. Situasjonen for desse ordbøkene har vore svært usikker i seinare tid, fordi Universitetet i Oslo har vedteke at det ikkje lenger vil ta ansvaret for språksamlingane som ordbøkene er knytte til, og heller ikkje for BOB, NOB og rettskrivingsbasane Norsk ordbank i særleg lang tid. Til no har Språkrådet saman med universitetet hatt ansvaret for å drifte og vedlikehalde ordbøkene og Ordbanken. Det er no avklart at Universitetet i Bergen vil overta språksamlingane frå 2016 (Språkrådet 2015). Det er svært viktig for heile den norske skulen å ha tilgang til ein gratis ressurs som utan atterhald opplyser om korleis norske ord skal skrivast, bøyast og brukast etter offisiell rettskriving.

Litteratur

Ordbøker

BOB (1986–) = *Bokmålsordboka*. Oslo: Kunnskapsforlaget. <nob-ordbok.uio.no/perl/ordbok.cgi> (oktober 2015).

NOB (1986–) = *Nynorskordboka*. Oslo: Det Norske Samlaget. <nob-ordbok.uio.no/perl/ordbok.cgi> (oktober 2015).

Annan litteratur

Bjørneset (2015): *LEXIN i Norge – Hva sier brukerne?* Innlegg på den 13. Konference om Leksikografi i Norden 2015. <nfi.ku.dk/nfl2015/fagligt-program/Bjoerneset.pdf> (januar 2016).

- Graedler, Anne-Line (2015). *Lærerstudenters bruk av digitale ordbøker i engelskundervisningen*. Innlegg på den 13. Konferanse om Leksikografi i Norden 2015. <nfi.ku.dk/nfl2015/fagligt-program/Graedler.pdf> (januar 2016).
- Karlsen, Knut E. & Dagfinn Rødningen (2008). Ordboksbruk i skolen – praksis og perspektiv. Utnytting av einspråklege ordbøker i norskfaget i den vidaregåande skolen. I: *LexicoNordica* 15, 93–114.
- Læreplan i norsk*. <www.udir.no/kl06/NOR1-05> (oktober 2015).
- Norsk ordbank. <www.edd.uio.no/prosjekt/ordbanken/> (januar 2016).
- Opplæringslova*. <lovdata.no/dokument/NL/lov/1998-07-17-61> (oktober 2015).
- Rambøll (2012). *Brukerundersøkelse av Lexin – ordbøker for minoritetsspråklige*. <www.udir.no/Tilstand/Forskning/Rapporter/Ramboll/Brukerundersokelse-av-LEXIN---ordboker-for-minoritetsspraklige/> (januar 2016).
- Språkrådet (2014): *Årsrapport 2014*. Oslo: Språkrådet. <www.sprakradet.no/globalassets/vi-og-vart/om-oss/armeldinger/arsrapport-2014-sprakradet.pdf> (januar 2016).
- Språkrådet (2015): *Språksamlingene til Bergen*. <www.sprakradet.no/Vi-og-vart/hva-skjer/Aktuelt/2015/spraksamlingene-til-bergen/> (januar 2016).
- TNS Gallup & Språkrådet (2014). *Digitale ordbøker i bruk. Undersøkelse blant elever og lærere på mellom- og ungdomstrinnet og i den videregående skolen*. <www.sprakradet.no/globalassets/sprakavare/rapporter/rapport---digitale-ordboker-i-bruk.pdf> (januar 2016).

Marit Hovdenak
seniorrådgjevar
Språkrådet
Postboks 8107 Dep
NO-0032 OSLO
marit.hovdenak@sprakradet.no

Ingunn Indrebø Ims
rådgjevar
Språkrådet
Postboks 8107 Dep
NO-0032 OSLO
ingunn.indrebo.ims@sprakradet.no

Ordbog over det norrøne prosasprog Online: struktur og brug

Ellert Þór Jóhannsson & Simonetta Battista

The topic of this article is the structure of *A Dictionary of Old Norse Prose* (ONP) and the use of its online version. This web-based dictionary reflects the complex history of the ONP dictionary project, which has evolved from a traditional print publication to an online version. In this article, we discuss the editorial approach to the material and demonstrate some of the features of ONP Online. We look at various aspects of the online dictionary and consider some of the benefits of this new medium.

1. Projektets baggrund

Ordbog over det norrøne prosasprog (ONP) er et ordbogsprojekt under Nordisk Forskningsinstitut ved Københavns Universitet, som registrerer ordforrådet i de oldnordiske prosatekster overleveret i norske og islandske håndskrifter fra midten af 1100-tallet til slutningen af middelalderen (for det poetiske sprog, se Finnur Jónsson 1931). Projektet startede tilbage i 1939, og i løbet af de første 40 år bestod arbejdet primært i excerpering af udgaver og håndskrifter, hvilket resulterede i et omfangsrigt arkiv bestående af i alt ca. 750.000 håndskrevne sedler. En oprindelig planlagt trykt publikation på 13 bind startede i 1989, men blev indstillet i 2004, efter at registerbindet (ONP Registre) plus tre ordbogsbind (ONP 1-3) var udkommet. Kort derefter gik man i gang med at digitalisere ordbogsmaterialet med det formål at gøre det tilgængeligt på internettet (for et detaljeret historisk overblik se Battista & Jóhannsson 2014).

2. ONP's tilgang og artikelstruktur

ONP har fra starten haft mange særlige træk, som adskiller den fra tidligere oldnordiske ordbøger. For at illustrere det nærmere kan vi se på en typisk ordbogsartikel fra den trykte udgave som fx *bátr* 'båd' i Figur 1.

bátr *sð. m.* [-s, dat. -i/-; -ar]

I. *bád, skibsbád* // *boat, ship's boat*: Þeir Þórðr fundu einn *bát*, er fiskimenn váru á *Stu^{II}R11127^x 33²⁴*; Laxa nya oc sva aðra nya fiska oc ostrur. þat scal kaupa a *batom* eða a bryggium ef uil. en eigi flytia i buðir til mangs *Bl 263⁵*; skaltu æigi þurfa að æiga biorg þina alla undir *báti* þessum *Róm^v 119⁸* (~*lat. carinae LucBell 140⁸* (*V 534*)); þangat scyldi færa *bát* nocqvern *Ridd623 31³¹* (~*lat. vehicula Milit 21b²⁰*); Þorgils mælti við hann, hvi hann færri svá óvarliga. Sturla svarar: lát mik sjá fyrir *báti* mínum (s: *lad mig sætte kursen* // s: *let me decide the course*), sem auðit má verða *Stu^{II}R11127^x 268¹*; Sægl oc akkeri baat oc ræiða allan. tiold oc allt þat sem til skips hoyrir *Bl 200¹²*; Þeir höfðu á skipinu litinn *bát* ok geck Egill þar á við þriðja mann *Eg 201¹⁷*; hliópu þeir til skipa ok hrundu fram *bátum* *MarB 94²⁷* (~*lat. litorea cimbas MirMar1898 24²³*); var þá engi *bátrinn*; því hann hafði brotit fyrir þeim *Trist^IX 196¹⁵* (~*fr. batel Tristan 172²⁰*); hann seldi þeim *bátinn* frá hafskipinu *Eb^x 77³*; Ingimundr var lagiðr í *bátinn* frá skipinu Stiganda ok búit um virðuliga sem þá var siðr um tigna menn

Vatn^x 55⁶; *item: •Ágr 24⁵*; *BlFar 281¹⁶*; *D11V (*1397>AM 260^x) 227¹*; *Eg 84⁵* **ausa bát sinn** (øse sin bád s:) "lade vandet", *urinere* // (*bail one's boat s:*) "make water", *urinate*: standit vpp með mer. ec vil ganga a borþ a/sa bat min *Mork 411²⁴*; þat uil eg, attu gefe mier tomstund til at ausa bat minn *Jes510 94¹⁶* **skjóta báti (útbyrðis), skjóta báti til** [!] *sætte skibsbád i vandet* // *put out a ship's boat*: er þau koma út at eyjum, bað hon þá skjóta *báti* útbyrðis, er stóð á ferjunni *Laxd 102⁵*; legiaz þeir um akri Sþan scutu þeir *báti* *ÓH 354⁶*; Þeir letu ut ok (urdu) aprtr-reaka undir Grims-ey; ok er þeir komv þar at, skuto þeir *bate* *Stu^IK 426¹¹*; þeir skvto bati til af skipi Þorbergs. ok geck stafninn or batinom *HákFris 431¹⁷*

II. (*cogn.*) Biorgolfr batr *BoglEirsp 464³⁶*; Fell Hallvarðr batr við tiotugu manna *AmH^x 74²*

Comp.: eptir-; fiski-; flatbytnu-; kaupskips-; knarrar-; sela-; skip-; smá-; smal-; uppgjafa-; vatn-

Gloss.: *EJ; CIV; Fr; Hertzþ; Fischer; Bin; Rim; LP; AJ; deVr; Fr4; NO; ÁBlM; (Bl)*

Litt.: *Vigeland 1957 [KLN M 2] 467-471*

Figur 1: Opslagsordet *bátr* 'båd' fra den trykte udgave af ONP (ONP 2:85-86).

ONP lægger megen vægt på at gengive flere eksempler og medtage yderligere oplysninger om hvert enkelt opslagsord, end man kendte fra tidligere værker. Det kan man tydeligt se, når man sammenligner opslaget fra ONP med den tilsvarende artikel fra en af ONP's forgængere, fx Johan Fritznors *Ordbog over Det gamle norske Sprog* (Fritzner 1886), eller *An Icelandic-English Dictionary* af Cleasby & Vigfusson (1874) (cf. Figur 2).

bátr, *m.* (G. -s) *Naab. Fm. VI, 446*; skjóta *báti* s: sætte *Skibsbaad* ud paa *Spæn, OH. 136⁹*; *Laxd. 30 (76⁹)*; *ausa bát sinn*, þom egentlig þeyðer at þse fin *Naab tom, bruges Fm. VII, 331*; *Mork. 205²⁹* om at lade fit *Naab*; þom *Þegenavn* Arni b. *DN. II, 68²⁵*; Þorsteinn b. *Flat. III, 226³⁶*; Høllvarðr b. *Flat. III, 548*; Björgolfr b. *Fm. IX, 39, 48.*

BÁTR, *m.* [a Scand. and Low Germ. word used in A. S., Engl., Dutch, but alien to O. H. G. and middle H. G.; even Luther (v. Grimm s.v.) never uses the word; it was later introduced into mod. High Germ., but has a foreign sound there. (Engl. *t* answers to High Germ. *z*); the word is in Germ. borrowed from Dutch or English] :—a *boat*, either a small open fishing vessel or a *ship-boat*. In Icel. only small boats are called so, those of two or four oars; an eight-oared boat is a 'ship.' *Eg. 121, 373, Eb. 142, Nj. 122, Jb. 398, Bs. i. 422, 423* in phrases, *ausa bát sinn*, *Fms. vii. 331*; *sjá fyrir báti sínum*, *to go one's own course, to mind one's own business*, *Sturl. iii. 247*; alliteration, *eiga byggð í báti*, metaph., *Bs. i. 422*. **COMPS:** *báts-borð*, *n. the side of a boat*, *Sturl. i. 119*. **báts-farmr**, *m. a boat's freight*, *Ann. 1342.*

Figur 2: Ordbogsartikler fra *Ordbog over det gamle norske sprog* (Fritzner 1886:117) og *An Icelandic-English Dictionary* (Cleasby & Vigfusson 1874:55).

Udgangspunktet i ONP's tilgang til ordstoffet er filologisk nøjagtighed og gengivelsen af citaternes ortografi og tekstforhold. Eksemplerne i ONP er ikke normaliseret, men viser stavemåden i det pågældende håndskrift eller i en kritisk udgave, og et detaljeret referencesystem (signaturer i kursiv) henviser til det aktuelle håndskrift, hvert citat er taget fra.

Nordiske Studier i Leksikografi 13 (2016): 165-175

Ordbogsartiklens øvrige oplysninger afspejler også ONP's detaljerede tilgang til ordstoffet. Her finder vi bl.a. henvisninger til udenlandske paralleltekster (især latinske) i de tilfælde, hvor den norrøne tekst er en oversættelse. Ud over dette er artiklen organiseret med henblik på syntaktiske træk, fx hvilke sætningsled der hører sammen, og fraser og kollokationer, især i forbindelse med verber. I slutningen af artiklen har man en del yderligere oplysninger, fx en liste over sammensatte ord, hvor opslagsordet udgør sidste led (*Comp*), en liste over tidligere ordbogsværker, hvor ordet forekommer (*Gloss*), samt relevante litteraturhenvisninger (*Litt*).

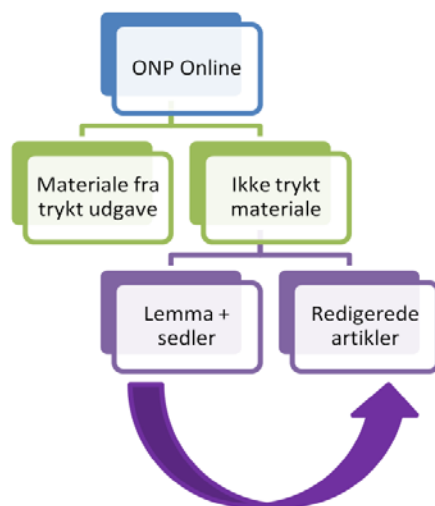
Andre oplysninger, som forekommer i selve artiklens opbygning, tager også udgangspunkt i loyalitet over for teksten. Derfor er morfologiske oplysninger (bøjningsmønster og verbal konjugation) baseret på belæg fra teksterne, dvs. bøjningsformerne skal være atteret i citaterne.

Ordbogen har to målsprog: dansk og engelsk. Definitionerne er skrevet på disse to sprog, selv om metasproget ellers er dansk. Dog består metasproget af mange (ofte latinske) forkortelser, som i de fleste tilfælde er de samme for begge sprog.

3. ONP Online – struktur og redaktionel arbejdstilgang

Omstillingen til digital udgave medførte helt fra starten en anden redaktionel tilgang. Målet med arbejdet var ikke længere at udgive fysiske bøger, men snarere at publicere ordbogsartikler gradvist på nettet og at udvikle ONP som et alsidigt leksikografisk værktøj. Samtidig skulle man bibeholde det videnskabelige niveau i behandlingen af ordstoffet og gengivelsen af detaljerede oplysninger. En væsentlig del af arbejdet har bestået i at integrere det allerede trykte materiale i den digitale version og i at udvikle redaktionelle metoder, som er tilpasset det nye medium (cf. Johannsson & Battista 2014).

Resultatet af dette arbejde har siden 2010 været tilgængeligt i den nye digitale udgave af ONP, som er udgivet på internettet, og som vi kalder ONP Online eller ONP netversion. Figur 3 giver et skematisk overblik over, hvordan onlineudgaven kombinerer både ordstof fra den tidligere trykte udgave og det ikke-trykte materiale.



Figur 3: Strukturen af ONP Online. Figuren viser de forskellige komponenter, der udgør onlineudgaven. Pilen skal vise, hvordan det materiale, som består af opslagsord/lemma+sedler, løbende bliver behandlet og erstattet af redigerede ordbogsartikler.

Da arbejdet med onlineudgaven først gik i gang, blev det besluttet at give brugerne adgang til ONP's hele materiale, dvs. både det, som allerede var udkommet på tryk, og det, som ikke havde været behandlet. Det ikke-trykte materiale blev gengivet i meget grov form som opslagsord sammen med tilhørende skannede citatsedler. Efterfølgende fortsatte redigeringsarbejdet dog, så nye redigerede ordbogsartikler kunne blive lagt direkte ud på nettet.

Overgangen til online digital udgave har betydet, at det redaktionelle arbejde sker på anden måde end tidligere. I den trykte udgave blev opslagsordene redigeret i stringent alfabetisk rækkefølge. I onlineudgaven redigeres opslagsordene efter en opdeling i tolv forskellige grupper baseret på ordklasse og morfologisk kompleksitet. Grupperne er: simpleks substantiver (< 10 citater), simpleks substantiver (≥ 10 citater), sammensatte substantiver, verber, simpleks adjektiver, sammensatte adjektiver, simpleks adverbier, sammensatte adverbier, pronomener, talord, konjunktioner og præpositioner. Denne opdeling har gjort redigeringsprocessen mere overskuelig og vist sig at være en nyttig måde at optimere det redaktionelle arbejde på.

Samtidig med opdeling af ordforrådet i forskellige grupper skelner vi mellem semantisk og strukturel redigering. For eksempel redigeres ikke-

sammensatte substantiver både semantisk og strukturelt, dvs. de struktureres efter definitioner og følger en traditionel ordbogsstruktur. Verber og præpositioner bliver i første omgang kun redigeret strukturelt, dvs. ikke semantisk, men snarere med henblik på form og syntaktiske træk. Her har man vurderet, at afklaringen af strukturelle komponenter ville være af betydelig hjælp for brugeren, selv om betydningen mangler. På længere sigt håber vi dog at kunne behandle alle ord både semantisk og strukturelt.

Andre ændringer i redigeringsproceduren inkluderer begrænsningen til kun et målsprog, dansk. Morfologisk information er ikke givet lige efter opslagsordet som i den trykte udgave. Vi viser heller ikke systematisk udenlandske paralleller i artiklen som før, men oftest kun på de skannede sedler. Til gengæld har vi øget antallet af fraser og kollokationer for at gøre ordbogsartikler med mange citater mere overskuelige.

Erfaringen med den ændrede redigeringsprocedure har været god. Vi har prioriteret redigering af præpositioner, verber og usammensatte substantiver. På nuværende tidspunkt (sommer 2015) har vi redigeret strukturelt og/eller semantisk ca. 57 % af ONP's materiale. Omkring 15 % var allerede omfattet af de trykte bind, mens 42 % kun er redigeret i onlineudgaven. Det betyder, at ca. 43 % af materialet endnu ikke er behandlet, men dette tal falder gradvis i takt med, at vi redigerer de resterende ordgrupper.

4. Brug af ONP Online

I dette afsnit vil vi give et overblik over ONP Online og de forskellige funktioner, som er frit tilgængelige på ONP's hjemmeside: <onp.ku.dk>. Her findes også en række andre informationer og supplerende oplysninger. Et screenshot fra hjemmesiden vises i Figur 4.

På hjemmesiden finder man genveje til forskellige undersider, bl.a. ONP's registre. Adgang til ordbogsmaterialet kan ske gennem basis eller avanceret ordliste. Vælger man en basis ordliste, får man opslagsordene i alfabetisk rækkefølge. Ved at vælge den avancerede ordliste får brugeren et søgevindue op, hvor man kan indtaste det ønskede ord eller den ønskede orddele (cf. Figur 5). Her har man muligheder for at skræddersy søgningen på forskellige måder.

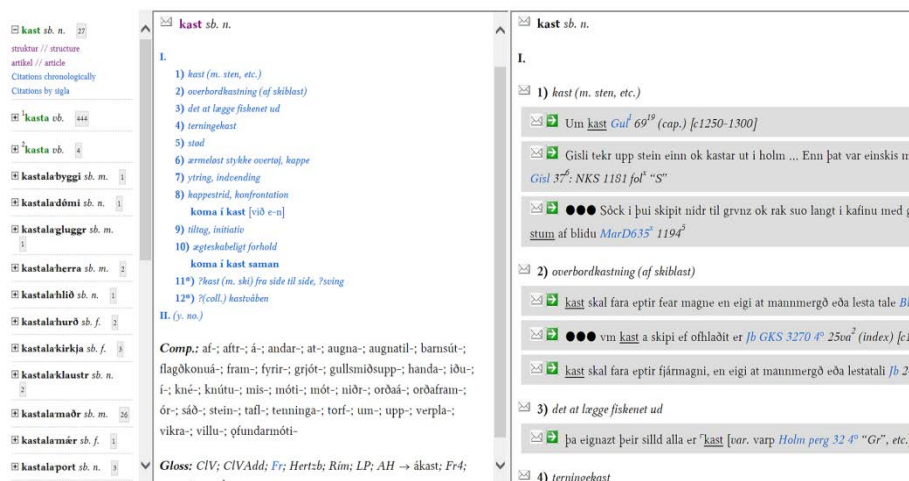


Figur 4: ONP's hjemmeside for ONP Online: *onp.ku.dk*. Adgang til Ordliste og Registre.



Figur 5: ONP's søgevindue i avanceret søgning. *Søgestreng*: Søg efter specifikt ord eller orddel, og brug særlige norrøne tegn ved at klikke på det pågældende symbol. *Søgemåde*: Vælg, om søgestreng skal være i begyndelsen, midten eller slutningen af et ord. *Ordklasse*: Begræns søgning til en bestemt ordklasse (f. eks. substantiver eller verber). *Grammatiske specifikationer*: Søg efter grammatiske træk som fx køn (mask., fem. eller neutr.).

I overensstemmelse med de forskellige komponenter, der udgør ONP (se Figur 3), findes der tre hovedtyper af opslagsord, nemlig opslagsord, som er redigeret i den trykte udgave, opslagsord, redigeret i onlineudgaven, og opslagsord, som endnu ikke er redigeret. Figur 6 viser resultatet af søgning på et typisk substantiv, *kast* 'kast', som hører til den gruppe ord, der har været redigeret i onlineudgaven.

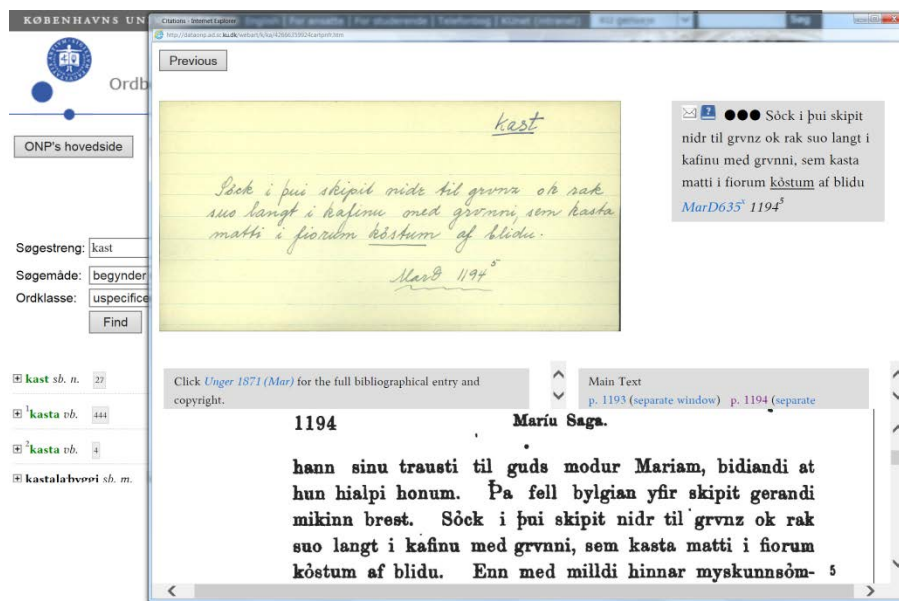


Figur 6: Screenshot fra ONP Online. Opslagsordet er *kast* 'kast'. Her vises hovedvinduet med ordliste, artikelstruktur og citater.

Ord, som begynder med *kast-*, vises på ordlisten til venstre. En separat spalte viser artiklens struktur i blå og giver et hurtigt overblik over dens opdeling i betydninger og fraser. Dette er især nyttigt, når det drejer sig om ord med mange definitioner og citater. *Comp.*, *Gloss* og *Litt* findes også her (cf. afsnit 2). Hvis man klikker på *Fr* i *Gloss*, får man det pågældende sted i en digital version af Fritznors ordbog (Fritzner Online) op. Foreløbig er det kun denne, der er linket til ONP's artikler på den måde.

Til højre vises den udfoldede artikel, dvs. selve citaterne under den relevante definition. De tre prikker ●●● foran nogle af citaterne indikerer et illustrativt eksempel eller særlig interessant citat. Man kan yderligere klikke på selve signaturen eller henvisningen, som står efter hvert eneste citat (i blå). Derved åbnes der et nyt vindue med adgang til ordbogens register,

hvor man kan se, præcis hvilket håndskrift der ligger til grund for den udgave, ordbogen henviser til, og citatet er taget fra.



Figur 7: Screenshot fra ONP Online. Pop-up-vindue med nærmere oplysninger om citat 3 under definition 1.

Ved at klikke på andre links åbnes der ekstra vinduer, hvor man kan finde endnu flere oplysninger. Når man klikker på den grønne pil ved et af citaterne, åbnes der et vindue med nærmere oplysninger om det enkelte citat (cf. Figur 7). Her får man et billede op af den skannede seddel og det indtastede citat. I samme vindue får man også et billede op af den side i udgaven, signaturen henviser til. Der er yderligere mulighed for større kontekst ved at blade en side frem og en side tilbage. Her kan man også klikke på udgavehenvisningen for at få bibliografiske oplysninger om udgaven frem.

Ved at klikke på en konvolut åbnes et vindue, hvor brugeren kan bringe forslag og kommentarer til ONP's medarbejdere. Her er en hurtig forbindelse med ordbogens stab, hvis man har forslag til forbedringer eller ændringer.

4.1. Fordele ved ONP Online

Selv om ONP Online ikke præcis indeholder alle de samme detaljerede informationer, man finder i de trykte ordbogsbind, har onlineversionen dog mange fordele for brugeren, når man sammenligner den med den trykte udgave:

- Fri adgang til hele materialet på nettet. Brugeren har adgang til alle skannede sedler og alle redigerede ordbogsartikler, både fra den trykte udgave og senere redaktioner.
- Mulighed for at læse citater i større kontekst. Brugeren har adgang til skannede tekster og udgaver.
- Overblik over artiklens struktur uden citater.
- Placering af alle citater under den pågældende definition i redigerede artikler.
- Flere søgemuligheder, som gør det muligt at skræddersy søgninger efter brugerens behov og interesser.
- Mulighed for at give feedback til redaktionen.
- Regelmæssig opdatering af ordbogen. Man får altid de nyeste oplysninger, hver gang man slår noget op.

Der er også nogle fordele for redaktørerne ved onlineudgaven:

- Mere produktiv redigeringsprocedure: mindre tid til korrektur-læsning og mere tid til redigering.
- Plads til større kontekst og yderligere oplysninger.
- Mulighed for løbende revisioner og tilføjelser, hvis nødvendigt, efter at artiklen er blevet publiceret på nettet.
- Feedback fra ordbogens brugere, som kan indgå i revision af artikler.

4.2. Fremtidsperspektiver

Vi har prioriteret nogle aspekter, som vi håber på at inkorporere i onlineudgaven inden alt for længe. Vi ser frem til igen at bruge engelsk konsekvent som målsprog ved siden af dansk. Vi kommer også til at udvide søgemulighederne gradvist. I den nuværende version kan man kun søge efter normali-

serede opslagsord. Databasen har dog flere oplysninger, som kan bruges til at tilføje søgemuligheder, som endnu ikke er en del af ONP Online. Fx søgning efter sideformer (unormaliserede former), som vi håber at kunne tilføje ved den næste store opdatering.

Længere ind i fremtiden har vi tænkt os at udvide mulighederne for at bruge ONP Online. For eksempel ville det være oplagt at kunne søge efter alle citater fra et bestemt værk eller begrænse søgningen efter perioder, geografisk område, håndskrifter, teksttype, etc. Desuden ville vi gerne gøre det muligt at søge i selve citaterne (tekstsøgning). Omfattende videnskabelige udgaver udgivet efter 2005 er heller ikke inkorporeret i ordbogens database endnu. Vores mål er at kunne holde ordbogen opdateret og tilføje henvisninger til de nyeste tilgængelige videnskabelige udgaver.

5. Konklusion

I denne artikel har vi diskuteret *Ordbog over det norrøne prosasprog*, dens karakteristiske træk og tilgang til ordstoffet, og hvordan de redaktionelle principper, som var udviklet til den trykte version, er blevet indarbejdet og videreudviklet for at tilpasse sig onlineudgaven. ONP Online er et dynamisk leksikografisk værktøj, som byder på mange spændende muligheder. Gennem det nye medium har brugerne fået forskellige nye måder at søge information og bruge ordbogens oplysninger på, ud over at have adgang til både redigeret og uredigeret materiale. Der er dog rum for udvidelser og forbedringer, og redaktørerne glæder sig til at arbejde videre med at udvikle onlineudgaven til gavn for alle dem, som er interesseret i de nordiske sprog i middelalderen og i nordisk litteratur og historie.

Litteratur

Ordbøger

Cleasby, Richard & Gudbrand Vigfusson (1874): *An Icelandic-English Dictionary*. Oxford: Clarendon Press.

Nordiske Studier i Leksikografi 13 (2016): 165-175

- Finnur Jónsson (1931): *Lexicon poeticum antiquæ linguæ Septentrionalis / Ordbog over det norsk-islandske skjaldesprog*. Opr. udg. af Sveinbjörn Egilsson. 2. udg. København: Kongelige nordiske oldskriftselskab.
- Fritzner, Johan (1886, 1891, 1896): *Ordbog over Det gamle norske Sprog* 1-3, rev. udg. Kristiania: Den norske forlagsforening.
- Fritzner Online. Johan Fritzners *Ordbog over Det gamle norske Sprog* <www.edd.uio.no/perl/search/search.cgi?appid=86&tabid=1275> (juni 2015).
- ONP = Helle Degnbol, Bent Chr. Jacobsen, James E. Knirk, Eva Rode, Christopher Sanders & Þorbjörg Helgadóttir (red.): *Ordbog over det norrøne prosasprog / A Dictionary of Old Norse Prose*. ONP Registre (1989). ONP 1: a-bam (1994). ONP 2: ban-da (2000). ONP 3: de-em (2004). København: Den Arnamagnæanske Kommission.
- ONP Online. *Ordbog over det norrøne prosasprog Online*. <onp.ku.dk> (juni 2015).

Anden litteratur

- Battista, Simonetta & Ellert Þór Jóhannsson (2014). Ordbog over det norrøne prosasprog 2004-2014 — fra trykt udgave til netversion. I: *LEDA-Nyt* 57, 6-23.
- Jóhannsson, Ellert Thor & Simonetta Battista (2014). A Dictionary of Old Norse Prose and its Users – Paper vs. Web-based Edition. I: Abel & al. (eds.): *Proceedings of the XVI EURALEX International Congress: The User in Focus, 15-19 July 2014*. Bolzano/Bozen, 169-179.

Ellert Þór Jóhannsson
ordbogsredaktør, ph.d.
Københavns Universitet
Njalsgade 136
DK-2300 København S
nk950@hum.ku.dk

Simonetta Battista
ordbogsredaktør, ph.d.
Københavns Universitet
Njalsgade 136
DK-2300 København S
sb@hum.ku.dk

Bruger- og ekspertinddragelse ved udarbejdelse af online (fag)ordbøger: det kooperative princip i leksikografien

Patrick Leroyer

This article outlines the tenets of a cooperative framework between expert-users, experts and lexicographers in the construction of not only specialised, online dictionaries but also of language for general purposes dictionaries. Firstly, the article briefly explains why the topic of cooperation with experts has not attracted the attention it deserves: this is due to the side effects of what might be called general language lexicographic hegemony, in which a strong consensual focus is awarded to general language dictionaries and to knowledge of language. Secondly, the article briefly refers to a few case studies of lexicographic cooperation – and of its counterpart, absence of cooperation. Finally, a cooperation model is presented on the basis of two closely related cooperation theses. It is concluded that lexicography could benefit from a greater interest in specialised lexicography in order to adequately meet the knowledge needs of dictionary users in the modern knowledge and information society.

1. Indledning – sproglig viden vs faglig viden

Der synes i leksikografien, inklusive den digitale leksikografi, at være en stærk tradition for at tale om *ordbogen* uden at præcisere ordbogstypen, men hvor der reelt set hovedsageligt tænkes på den almensproglige ordbog. I forlængelse heraf synes der også at være tradition for at betragte ordbogen som kilde til *sproglig viden* i bred forstand. Indkaldelsen til den 13. Konference om Leksikografi i Norden er ingen undtagelse:

Vigtigheden af at interessere sig for, hvordan *sproglig viden* opsøges, tilegnes og bruges på digitale medier, understreges fx af den stærkt øgede brug af nye platforme [...] (NFL 2015, min understregning)

Ordbogen som kilde til *faglig viden* forbigås i tavshed. Dette skyldes sikkert, at fagordbogen¹ tilsyneladende ikke anses for at være i stand til at kunne mønstre lige så mange potentielle brugere som den almensproglige ordbog. Fagordbogens betydning for informations- og videnssamfundet bør dog ikke undervurderes. Fokus på ordbogstypen har stor betydning for leksikografien. Det kan antages, at hvad der synes at være rigtigt og smart for udarbejdelsen af den almensproglige ordbog, der blandt andet bygger på sprogvidenskab, ikke nødvendigvis er det for fagordbogen, der blandt andet bygger på andre videnskaber, herunder inddragelse af brugerne, inklusive ekspertbrugere, og fageksperter i ordbogsprocessen, jf. Klosa 2013, som i sin leksikografiske model over ordbogsprocesser blot henviser til *dictionaries in general*. Samarbejdet med brugerne og fageksperter i ordbogsprocesserne undervurderes også, selvom der er behov for kooperativ leksikografi.

I dette bidrag vil jeg forsvare to leksikografiske teser om ordbogsprocesser og leksikografisk samarbejde, idet jeg vil trække på erfaringerne fra opbygningen af to online fagordbøger, vinordbogen Oenolex 2015 (Leroyer 2013a og 2013b; Leroyer & Høy 2014) og Ejendomsordbogen 2015 (Leroyer & Hansen 2012). Begge teser forsvare samarbejdet med ekspertbrugere og fageksperter i ordbogsprocesserne, både ved online fagordbøger, men også ved almensproglige ordbøger. Jeg vil først komme ind på den leksikografiske diskrimination af fagordbogen og søge at give et bud på dens årsager. Jeg vil i denne forbindelse bringe nyere talstatistik, der indikerer, at fagordbogen trods den leksikografiske diskrimination og synes at nyde stor interesse blandt brugerne. Dette rokker ved almenordbogens privilegerede position. Jeg vil efterfølgende illustrere samarbejdet med brugere og fageksperter gennem en kort omtale af Oenolex 2015 og Ejendomsordbogen 2015. Jeg vil også inddrage en almen (fag)-ordbog til oversættelse, der tydeligvis udviser tegn på manglende leksikografisk samarbejde med fageksperter, hvilket i sidste kan ende med at kompromittere dens oversættelsesfunktioner. Til sidst vil jeg opstille det kooperative princip i leksikografi i form af to samarbejdsteser, der bygger på tilsvarende samarbejdsmodeller (efter Klosa 2013).

¹ I dette bidrag bruger jeg termen *fagordbog* som generisk betegnelse for leksikografiske opslagsværker, hvori adgang til viden om fagsprog og/eller faglig viden er i fokus. De forskellige betegnelser (fagordbog, leksikon, encyklopædi, glossar) refererer til samme leksikografiske virkelighed, cf. Olivera & Tarp (2014).

2. Ordbogen som kilde til sproglig – og til faglig viden

2.1. Almenordbogens hegemoni

At ordbogen blandt leksikografer primært ses som kilde til *sproglig viden* i bred forstand kan betegnes som et udslag af almenordbogens hegemoni, hvorved almenordbogen uforholdsmæssigt kommer til at dominere store dele af den leksikografiske teori og praksis og derved overskygge andre typer af ordbøger. Dette hegemoni skyldes givetvis følgevirkningerne af en klar sprogvidenskabelig præference i leksikografien, der blandt andet fører til en definatorisk uklarhed om, hvad vi som leksikografer egentlig lægger i begrebet *ordbog*. Konsekvenserne for dette hegemoni er blandt andet, at betydningen af den nødvendige interdisciplinaritet i leksikografien undervurderes, hvilket rammer samarbejdet med brugere og eksperter om ordbogsprojekter uanset type. At ordbogens potentiale som væsentlig kilde til *faglig viden* undervurderes, er også u hensigtsmæssigt i informations- og videnssamfundet, for dermed mister leksikografien en stor del af sine udviklingsmuligheder i en tid, hvor de store nationale ordbogsprojekter er truet af støt faldende bevillinger og mindre politisk bevågenhed. Det store fælleseuropæiske samarbejde om e-leksikografien i ENeL (2015) netværket under EU-COST-programmet er et forsøg på at modvirke denne udvikling, men det er bekymrende, at fagleksikografien og fagordbøger synes at være mere eller mindre udelukket fra dette samarbejde.

Betyder det så, at brugerne rundt om i Europa, og i verden for den sags skyld, ikke interesserer sig synderligt for fagordbøger? Kan man kvantificere brugernes interesse²? Ved en sammentælling af titlerne på nye ordbogsudgivelser blev det i Leroyer (2011) påvist, at fagordbøger udgør 75 % af nye ordbogsudgivelser, mens almensproglige ordbøger kun tegner sig for de resterende 25 %. Hvor disse tal vidner om klare og veldefinerede vidensorienterede informationsbehov i samfundet og i øvrigt bekræfter ordbogens store potentiale som informationskilde, siger de naturligvis ikke ret meget om den reelle interesse blandt brugerne. Der kunne sagtens tænkes at eksistere en stor diskrepans på ordbogsmarkedet mellem fagordbogs-

² En kvantitativ vurdering vil kunne aflæses af salgstallene for ordbogsudgivelser, men disse er desværre ikke tilgængelige.

forfattere og deres potentielle brugere. Kilgariff (2012: 27) undrer sig over tallene, men gør med rette sine læsere opmærksom på denne mulige diskrepans, idet han sammenligner fagordbøger og almensproglige ordbøger ved hjælp af en lufthavnsmetafor. Fagordbøger skal her ses som de talrige lokale flyvepladser, mens almensproglige ordbøger skal ses som de langt færre, men betydeligt større internationale lufthavne, der selvsagt genererer betydeligt mere trafik og omsætning.

Det skal naturligvis ikke betvivles, at almensproglige ordbøger er centrale for leksikografien, og ej heller at almensproglige ordbøger nyder betydelig anerkendelse eller har en stor kulturrolle. For dette er netop min pointe: Kilgariff tildeler almensproglige ordbøger en række træk, der bestyrker deres hegemoni, vanens magt, indtjening og deres fysiske omfang, men undlader at henvise til kulturrollen eller til andre, mere saglige faktorer³. De store akademiske og videnskabelige almensproglige ordbøger vil jo typisk fremstå som væsentlige kulturbærere af – og kulturstøtter til den nationale identitet.

2.2. Udfordring af hegemoniet: ordbogen som kilde til faglig viden

Kigger man imidlertid på den globale trafik, der genereres på de store online encyklopædier, fx Wikipedia (Wiki 2015) med mange millioner artikler på mange sprog, og på online fagordbøger, leksika, fagglossarer og ordlister med flere, vil man hurtigt opdage, at den almene ordbogs hegemoni, således som Kilgariff fremstiller den, er truet, når vi inddrager det for leksikografien mest centrale parameter: brugernes faktiske brug af ordbøger. I Danmark blev der for eksempel i marts 2015 alene registreret 40.000 sidevisninger i timen, hvilket samlet set giver 960.000 daglige unikke sidevisninger eller tæt på hen ved en million unikke sidevisninger om dagen. Dette tal skal sammenholdes med de 60.000 daglige brugere, som Ordnet ifølge Magisterbladet af 17. januar 2014 får besøg af. Meget tyder således på, at interessen for at tilegne sig faglig viden er omfattende, og at fagleksikografien, skønt mindre prestigefyldt, spiller en væsentlig rolle i videns- og informationsområdet. Frygten for diskrepans på markedet kan manes til jorden.

³ Det er værd at bemærke, at Kilgariff ikke henviser til sprogvidenskabens dominans eller til kultur- og sprogpolitikken som mulige grunde for almenordbogens hegemoni, men udelukkende til en række *common sense* faktorer, som for eksempel opfattelsen af, hvad en ordbog er blandt *almindelige mennesker*.

2.3. Hvad mener vi egentlig med (fag)ordbog?

En del af forklaringen på almenordbogens hegemoni skal søges i den udbredte uklarhed, der hersker omkring defineringen af leksikografiens centrale genstand, selve *ordbogen* og de forskellige typer af ordbøger. Slår man op i DDO (2015) for at finde en definition på begrebet *fagordbog*, vil man søge forgæves, da termen ikke er medtaget.

fagordbog = [Der er ingen resultater med "fagordbog"]

Begrebet *ordbog* findes til gengæld og defineres meget lig den almen-sproglige ordbog, idet definitionen ser helt bort fra *faglige* eller fagsproglige oplysninger:

ordbog = opslagsværk der indeholder oplysninger om ord og deres sproglige form og indhold, fx stavning, ordklasse, bøjning, udtale, betydning, brug, synonymer og historie i et sprog eller om ordenes oversættelse til et andet sprog. Værket kan publiceres som en bog eller digitalt som fx en hjemmeside eller app. (DDO 2015)

Slår man op på *leksikografi*, finder man en kortfattet definition, der helt i tråd med det almensproglige ordbogshegemoni opfatter leksikografien som en sprogvidenskabelig disciplin. Mere overraskende er det dog at erfare, at definitionen af disciplinen også indbefatter udarbejdelse af *leksika*, der netop defineres ud fra deres faglige forankring. Dette virker noget inkonsekvent, og man kunne derfor frygte, at manglen på definatorisk stringens omkring selve leksikografiens genstand kan have negative konsekvenser for udviklingen af den leksikografiske teori og praksis, først og fremmest en manglende anerkendelse af, at der skal andre eksperter til end sprogvidenskabelige eksperter for at sikre høj faglig integritet og autoritet i udarbejdelsen af ordbøger, der i mindre eller større grad behandler faglige videnselementer, uanset længden på artiklerne i øvrigt.

leksikografi = gren inden for sprogvidenskaben der beskæftiger sig med udarbejdelse af ordbøger og leksika (DDO 2015)

leksikon = opslagsværk med de væsentligste oplysninger på ét eller flere bestemte fagområder eller på alle fagområder især i form af et bogværk med (kortfattede) artikler ordnet alfabetisk efter stikord (DDO 2015)

Som allerede nævnt vil jeg i den følgende del af artiklen læne mig op ad Olivera og Tarps (2014) definition, der etablerer, at fagordbøger blot er en fælles betegnelse for en lang række leksikografiske betegnelser, der alle har til fælles, at de bringer informationer, der vedrører fagområder, der falder uden for almen viden og kultur (Olivera & Tarp 2014)⁴.

3. Eksempler på leksikografisk samarbejde

3.1. Ejendomsordbog

Fast ejendom trækker på flere fagligheder herunder økonomi og finans, jura, arkitektur og byggeteknik. Ejendomsordbogen (2015) er en fransk-dansk brancheordbog (Leroyer & Hansen 2012), som ikke kunne være blevet til uden et omfattende samarbejde mellem en række ekspertbrugere og fageksperter, der samtidig også er ordbogsbrugere. Disse omfatter eksperter i ejendomshandel, i ejendomsfagsprog (ordbogsredaktionen) og i fagleksikografi (medarbejdere ved Center for leksikografi på Aarhus Universitet). Eksperterne deltager i alle faser af ordbogsprocesser, da ordbogen løbende udvides og opdateres i samråd med dem. I skrivende stund udvides ordbogen med engelske oversættelser til brug ved ESPI, den internationale ejendomsbrancheskole i Paris. Definitionerne på jobtyperne er fx forfattet af underviserne, og matcher derved kompetencemål for skolens uddannelser:

gestionnaire d'actifs immobiliers

Il élabore des stratégies de valorisation patrimoniales, conformément à la politique définie par la société et ses actionnaires, en mobilisant

⁴ Dermed ikke være sagt at udarbejdelsen af ordbogsartikler ved 'almene kulturord' og 'almen viden' ikke også kan kræve inddragelse af fageksperter. Grænse-udmålingen mellem alment sprog og fagsprog er glidende og stiller krav til ordbogsredaktioner, hvor ordbogen fx kan have dokumentationsformål.

l'ensemble des directions impliquées dans cet objectif de rentabilisation. Activités principales: stratégie patrimoniale, comptabilité et finance, mise en place et évolution des outils d'asset management, trésorerie.

3.2. Vinsmagningsordbog

Vinsmagningsordbogen *Oenolex 2015* (Leroyer 2013a og 2013b; Leroyer & Høy 2014) illustrerer ligeledes behovet for tæt samarbejde mellem leksikografer, fageksperter og ekspertbrugere. Disse omfatter vinsmagningslærere ved vinskolen, oenologer, undervisere, producenter, samt andre branchefolk. Som ved *Ejendomsordbogen* er eksperterne samtidig ordbogsbrugere. Ordbogskonceptet er bestilt af brugerne selv ud fra deres informationsbehov og efterfølgende tilpasset af leksikograferne. De leksikografiske data er genereret af brugerne, idet ordbogen bygger på et korpus af autentiske verbale, faglige interaktioner i forskellige vinsmagningssituationer. Det er også eksperterne, der udarbejder definitioner på de begreber, de anvender i forskellige sammenhæng, da formålet har været at give professionen nem og hurtig adgang til den diskursive dynamik, der præger erhvervet. Dette illustreres her ved definitionen af det strategiske AOC-begreb, hvori hovedargumentet ligger på selve kvalitetssikringen:

L'Appellation d'Origine Contrôlée (AOC) est une certification. Elle garantit les caractéristiques du vin: son terroir d'origine, son mode de production, qui suit des étapes bien précises, les savoir-faire locaux, inscrit dans la tradition, enrichie au fil du temps. Chaque vin estampillé AOC est soigneusement contrôlé, à chaque stade de son élaboration.

3.3. Den hybride ordbogs begrænsninger

Til sidst vil jeg bringe et eksempel på manglende samarbejde mellem leksikografer og fageksperter. Blinkenberg & Høybye (2015) (B&H), som er den største ordbog mellem dansk og fransk, er kommet online på *Ordbogen.com*. Det er karakteristisk for denne oversættelsesordbog, at den på én gang er almensproglig ordbog og fagordbog, altså en slags hybrid

ordbog. Det fremgår at ordbogens forord, at den medtager de vigtigste former for fagsprog, således også juridisk fagsprog fra den Europæiske Union samt fransk ret og retsvæsen. I Leroyer & Rasmussen (2010) analyseres en specifik oversættelsesproblematik i forbindelse med oversættelse af en EU-tekst, der afslører grundlæggende mangler ved de konventionelle fagordbøger og også ved B&H. Problematikken illustreres ved nedenstående eksempel, hvor tekst 1 er den originale EU udgangstekst, og tekst 2 forfatterens oversættelsesforslag. Problemet ligger i oversættelsen af præposition *sans préjudice de*, der bruges med vidt forskellige betydninger.

- 1) *Sans préjudice* des dispositions prévues aux articles 7 et 11, si la marchandise n'a pas reçu la destination prescrite, le montant des droits non perçus doit être payé aux autorités compétentes, *sans préjudice* des intérêts moratoires éventuellement exigibles.
- 2) Hvis varerne ikke er blevet anvendt i overensstemmelse med det foreskrevne formål, skal den indbetalte told *med tillæg af* eventuelt forfaldne morarenter betales til de kompetente myndigheder, *jf. dog* bestemmelserne i artikel 7 og 11.

Det fremgår af oplysningerne nedenfor, at B&H er ude af stand til at yde den fornødne hjælp til fagoversætteren:

- **sans** préjudice de a) uden skade for, uden at tage noget fra, uden at give afkald på b) uden hensyn til, uden at dermed noget præjudiceres med hensyn til uden at foregribe (n.) c) med forbehold af
- sans préjudice des dispositions de l'article 3 (dr) uden at bestemmelserne i paragraf 3 foregribes
- sans préjudice de mes droits uden at min ret trædes for nær
- sans préjudice du droit (el. de la faculté) de uden at retten til... afskæres (el. fortabes)
- sans préjudice du fait que uden hensyn til den omstændighed at

De angivne ækvivalenter, hvoraf kun én har fagmarkering *dr* (droit/juridisk), kan ikke bruges her, og ordbogen undlader desuden at bringe vigtige oplysninger om de mikrostrategier, der skal anvendes ud fra kontekst for at kunne

nå frem til den korrekte oversættelse. I kontrast til B&H's oplysninger vises nedenfor artikelforslag (Leroyer & Rasmussen 2010), der bygger på en korpusanalyse af danske og franske tekstbelæg for brugen af *sans préjudice de*. Løsningen er møntet ud på fagoversætterens behov og rummer en række mikrostrategier, der tager højde for de diskursive kontekster, hvori præpositionen anvendes ved forskellige argumentationsstrategier.

Sans préjudice de

1. Efterfulgt af lov, forordning, paragrafhenvi­sing
 - jf. dog (dog ikke først i sætning)
 - med forbehold
 - for så vidt andet ikke er fastsat
2. Ikke efterfulgt af lov, forordning, paragrafhenvi­sing
 - uden at det dog berører, idet det ikke berører, uden at det berøres
 - ved inanimeret logisk subjekt = uden at give afkald på
 - ved fremtidige relationer = uden at dette dog begrænser muligheden for, uden at dette er til hindring for
3. I faste udtryk
 - intérêts (interesse) = uden at tilsidesætte interesser
 - intérêts (renter) = med tillæg af renter
 - principe = uden at tilsidesætte princippet
 - sauf preuve du contraire = medmindre det modsatte bevises

Det skal gerne medgives, at kritikken af B&H online for manglende ekspertinddragelse i korpusstøttet juridisk ækvivalens måske ikke synes at være helt retfærdig, da ordbogen bygger på de trykte ordbøger, hvis udarbejdelse netop ikke var korpusbaseret. Ligeledes kan man forvente, som redaktørerne er inde på, at indholdet vil blive opdateret. Oversættelsesproblematikken peger dog på, at jura og juridisk oversættelse stiller store krav til tosprogsleksikografien. Viden om lingvistik og fagsprogslingvistik er ikke nok i sig selv og bør suppleres af den faglige viden, der skal sikre den leksikografiske autoritet. Problematikken peger også på, at almensproglige ordbøger tit og ofte, ligesom B&H konsekvent gør det, inddrager en lang

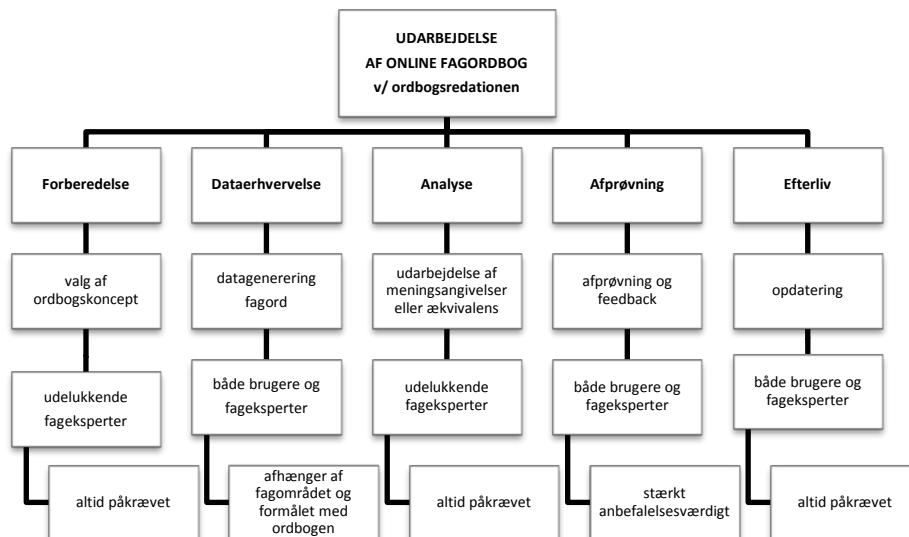
række fagord fra forskellige fagområder, hvorved medvirken af fageksperter trænger sig på, ikke mindst ved udarbejdelsen af definitioner og ved fastsættelsen af ækvivalens. Dette fører til opstillingen af to leksikografiske grundteser. I begge tilfælde er det ordbogsredaktionen, der har det overordnede ansvar under de forskellige ordbogsprocesser.

4. To leksikografiske grundteser

4.1. Udarbejdelse af online fagordbøger

I udarbejdelsen af online fagordbøger, hvor ordbogen er et værktøj for en virksomhed eller brancheorganisation, er inddragelse af både brugere og fageksperter påkrævet, da der er sammenfald mellem de to grupper.

- Ved udarbejdelsen af alle andre typer online fagordbøger er samarbejdet med fageksperter påkrævet i næsten alle faser af ordbogsprocessen.
- Som ved online almenordbøger er samarbejde med brugerne påkrævet under hele ordbogens levetid i ordbogens 'efterliv'.

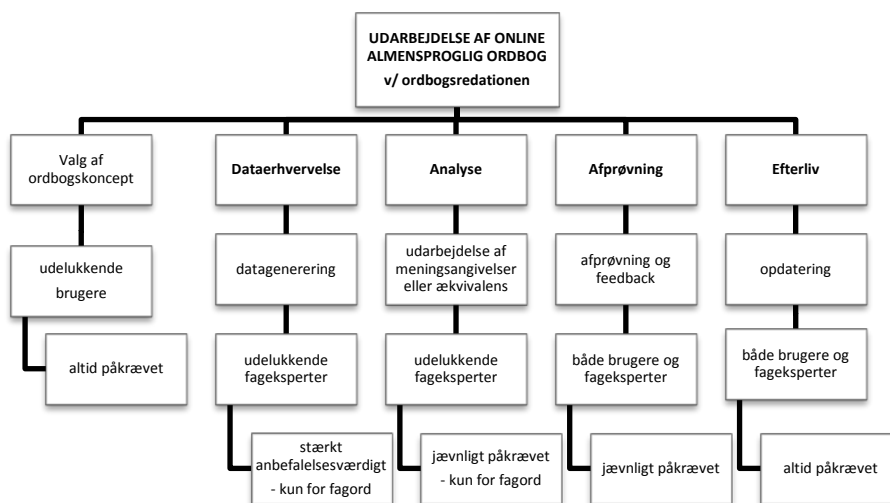


Figur 1: Samarbejde ved udarbejdelse af online fagordbøger.

4.2. Udarbejdelse af online almensproglige ordbøger

I udarbejdelsen af online almensproglige ordbøger er inddragelse af fagekspertter stærkt anbefalelsesværdigt, da næsten alle ordbøger vil inkludere en del fagord; visse ordbøger vil endda inkludere en større andel fagord, som fx Blinkenberg & Høybye-ordbøgerne (Schøsler 2014).

- Samarbejdet med brugerne er påkrævet i valget af koncept, da afklaringen af brugernes informationsbehov er afgørende for ordbogens kvalitet.
- Samarbejdet med fagekspertter er stærkt anbefalelsesværdigt i analysefasen, da ordbøgerne typisk inkluderer fagord. Dette er med til at sikre korrekte meningsangivelser og ækvivalensløsninger.
- Samarbejde med brugere og fagekspertter er stærkt anbefalelsesværdigt i ordbogens afprøvningsfase. Da online almenordbøger skifter status fra leksikografiske produkter til leksikografisk service bør samarbejdet være i fokus under hele ordbogens levetid i ordbogens 'efterliv'.



Figur 2: Samarbejde ved udarbejdelse af online almensproglige ordbøger.

5. Konklusion

Vi har i leksikografi opbygget en stærk lingvistisk tradition og præference, der har domineret de leksikografiske fagmiljøer siden lingvistikken opstod i forrige århundrede. Ekspertisen ligger i arbejdet med almensproglige ordbøger, i de store nationale ordbøger, i udviklingen af specielle sprogordbøger (f.eks. historiske eller dialektordbøger) og i sammenkædningen af eksisterende ordbøger i ordbogsportaler. Leksikografien kan også bidrage ved at tilknytte ordbogsressourcer til *linked data* og *digital humanities*. Men det moderne informationssamfund vil ikke kun have informationer om sprog i almensproglige ordbøger, hvor fokus normalt ligger på sproglig viden og hjælp til kommunikation. Det moderne samfund vil også have informationer om sagen og om verden gennem fagordbøger, hvor fokus normalt ligger på faglig viden og hjælp til videnstilegnelse. Efterspørgslen efter denne type information er enorm, mens udbuddet af kvalitetsfagordbøger er forsvindende lille. Det er derfor paradoksalt, at leksikografien, ved at holde den almensproglige, akademiske tradition i hævd, afskærer sig fra at udvikle det, den er bedst til: at koncipere leksikografiske informationsværkøjer, der bredt kan tilfredsstille de nutidige informationsbehov. Ved at undervurdere betydningen af fagordbogen som et centralt element i sit videnskabelige felt gør leksikografien sig en bjørnetjeneste. Den undergraver sin egen berettigelse og overlader markedet til Google og til andre aktører, der mangler en grundlæggende leksikografisk forståelse. Det vil i fremtiden blive stadig sværere for leksikografien at beskytte sin videnskabelige position og sin markedsposition, hvis den for ensidigt fokuserer på sin almensproglige ordbogsvirksomhed, som primært har til formål at pleje de store nationale ordbøger og deres fortsatte udvikling og udbygning som uundværlige sprogpoltiske instrumenter og kulturprodukter, der er medvirkende til at sikre nationens identitet. Dette er særdeles vigtigt, men politikkerne og det private erhvervsliv vil også være lydhøre over for støtte til og samarbejde med leksikografiske miljøer, der kan udvikle stærke vidensløsninger til de stadigt stigende vidensbehov i det højmoderne videns- og informationssamfund. Det er derfor bydende nødvendigt at satse på det kooperative princip i leksikografien og udbygge samarbejdet med ekspertbrugere og fageksperter for ad den vej søge at løse det leksikografiske paradoks, inden det er for sent.

Litteratur

Ordbøger og andre kilder

- B&H = Blinkenberg & Høybye, fransk-dansk: <www.ordbogen.com.ez.statsbiblioteket.dk:2048/opslag.php?word=pr%C3%A9judice&dict=bhfd> (september 2015).
- DDO = *Den Danske Ordbog*. <ordnet.dk/ddo> (september 2015).
- Ejendomsordbogen = Liselotte Kruse, Patrick Leroyer: *Ejendomsordbogen*. Ordbogen.com. <www.ordbogen.com> (september 2015).
- Magisterbladet* (2014) = <magisterbladet.dk/magisterbladet/2014/012014/012014_p8> (september 2015).
- ENeL = *European Network of e-Lexicography*. <www.elexicography.eu> (september 2014)
- NFL = Velkomsttekst i rundskrivelsen til 13. konference om leksikografi i Norden 2015 <nfi.ku.dk/nfl2015/om-konferencen/> (september 2015).
- Oenolex = Patrick Leroyer, Laurent Gautier, Hédi Maazaoui & Arnaud Millereux. *Oenolex*. Université de Bourgogne: Centre de ressources numériques <pandor.u-bourgogne.fr/Oenolex> (september 2015).
- Wiki = *Wikipedia*. Multilingual statistics. <stats.wikimedia.org/EN/Sitemap.htm> (juni 2015).

Anden litteratur

- Kilgariff, Adam (2012): Review of Pedro A. Fuertes-Olivera and Henning Bergenholtz (eds.), *e-Lexicography: The Internet, Digital Initiatives and Lexicography*. I: *Kernermann Dictionary News* July 2012, 26-29.
- Klosa, Annette (2013): The lexicographical process (with special focus on online dictionaries). I: Rufus Gouws, Ulrich Heid, Wolfgang Schweickard & Herberst Ernst Wiegand (Hrsg.): *Dictionaries. An international Encyclopedia of Lexicography. Supplement Volume: Recent Developments with Focus on Electronic and Computational Lexicography*. Berlin, Boston: de Gruyter, 517-524.
- Leroyer, Patrick (2011): Change of Paradigm in Lexicography: From linguistics to Information Science and from Dictionaries to Lexicographic Information Tools. I: Pedro A. Fuertes Olivera & Henning Bergenholtz (eds.): *e-Lexicography: The Internet, Digital Initiatives and Lexicography*. London: Bloomsbury Academic, 121-140.

- Leroyer, Patrick (2013a): New Proposals for the Design of Integrated Online Wine Industry Dictionaries. I: *Lexikos* 23, 209-227.
- Leroyer, Patrick (2013b). Putting words on wine: OENOLEX Burgundy, new directions in wine lexicography. I: Deny A. Kwary, Nur Wulan & Lilla Musyahda (eds.): *Lexicography and Dictionaries in the Information Age: Selected papers from the 8th ASIALEX International Conference*. Airlangga: Airlangga University Press, 228-235.
- Leroyer, Patrick & Liselotte Kruse Hansen (2012): Ejendomsordbogen fransk/dansk: ny integreret e-ordbog. I: Birgit Eaker, Lennart Larsson & Anki Mattisson (red.): *Nordiska studier i lexicografien 11. Rapport från Konferensen om lexicografi i Norden, Lund 24.-27. maj 2011*, 405-417.
- Leroyer, Patrick & Asta Høy (2014): OENOLEX Bourgogne: Lær at sætte ord på vin. Nyt vinordbogs-koncept, nye veje for brancheordbøger. I: Ruth E. Vatvedt Fjeld & Marit Hovdenak (red.): *Nordiska Studier i lexicografi, 12. Rapport fra Konferanse om lexicografi i Norden 13.-16. august 2013*, 287-302.
- Leroyer, Patrick & Kirsten Wølch Rasmussen (2010): Accessing Discursive Data Types in Legal Translating Dictionaries: The Case of Sans Préjudice de. I: D. S. Giannoni & C. Frade (eds.): *Researching Language and the Law: Textual Features and Translation Issues*, Vol. 121. Bern: Peter Lang, 157-178.
- Olivera, Pedro Fuertes & Sven Tarp (2014): *Theory and Practice of Specialised Online Dictionaries. Lexicography and Terminography*. Lexicographica Series Maior 146. Berlin/Boston: De Gruyter.
- Schøsler, Lene (2014): Blinkenberg og Høybyes Dansk-Fransk og Fransk-Dansk Ordbog – fra seddelsamling til trykte ordbøger og online udgave. I: *LexicoNordica* 21, 161-180.

Patrick Leroyer
lektor, ph.d.
Aarhus Universitet, Center for lexicografi
Jens Chr. Skous Vej 4
DK-8000 Aarhus C
pl@asb.dk

Den Danske Ordbog som app – hvorfor og hvordan?

Henrik Lorentzen, Nicolai Hartvig Sørensen & Lars Trap-Jensen

In addition to the existing online version, The Danish Dictionary was published as an app in May 2013. This paper sums up the considerations of the team who created the contents and design of the app. It attempts to explain why the app came to look the way it did by examining e.g. how the technical platform affects usage, how navigation can be optimized on a device with no physical keyboard, and what functionalities should be given priority in the app. The relatively small screen size of a handheld device, as compared to a laptop or desktop computer, forced the team to prioritize and include only the essential content elements. Financial constraints also played a role, and the paper discusses the consequences of a tight budget. Finally, the paper takes a look at other apps for future improvements, and considers the recommendations given in a critical review that followed.

1. Indledning

Den Danske Ordbog (DDO) udkom i appversioner til Android og iOS i maj 2013. Det er på den ene side ikke længere siden end at redaktionen har i frisk erindring hvilke idéer og overvejelser den gjorde sig i den forbindelse, og på den anden side er det tilstrækkelig længe til at redaktionen har kunnet gøre sig erfaringer og få tilbagemeldinger fra både brugere og kritikere. I denne artikel vil vi redegøre for hvilke overvejelser vi gjorde os før vi gik i gang, og hvorfor appen kom til at se ud som den gør. Vi ser på mulig indflydelse fra den teknologiske platform, beskriver vores designovervejelser og hvordan vi har forsøgt at sikre brugerne den bedst mulige oplevelse med at navigere på en enhed med en væsentligt mindre skærm end på en almindelig computer og uden et fysisk tastatur, herunder hvilke funktionaliteter der er blevet prioriteret i appen. Vi kommer ind på de begrænsede resursemæssige forhold der gjorde sig gældende under udviklingen af appen, og hvilke konsekvenser – ikke alle negative – de havde på valgene.

Desuden sammenligner vi kort appen med andre ordbogsapper og gennemgår forskellene. Hvor har vi endnu mangler? Hvor har vi løst problemer anderledes? Til sidst gennemgår vi hvordan appen blev modtaget, og hvilken kritik den har fået.

2. Hvorfor udgive en onlineordbog som app?

Den teknologiske udvikling har gjort at behovet for at konsultere opslagsværker har ændret sig. Det foregår ikke altid på kontoret eller i hjemmet foran en stor computerskærm. Vi har i dag adgang til den digitale omverden stort set altid, og når vi ikke er på arbejde eller på hjemmekontoret, sker kontakten især ved hjælp af små bærbare apparater, mobiltelefoner og tabletcomputere, som vi kan have med overalt. Derfor har brugerne en forventning om at de også kan komme i kontakt med vores ordbøger på den måde. For tiden sker det overvejende i form af små dedikerede programmer, apper til mobile enheder. Alternativet til appen ville have været at udvikle enten en særlig visning af hjemmesiden til små skærme eller hybridformen webapplikation. Den tilpassede mobilvisning anser vi da også for at være ønskelig, men i situationen tillod resurserne kun at prioritere den ene af mulighederne, og da vejede tilgængeligheden ved en egentlig, såkaldt "native" app tungest: det at brugerne kan søge og finde DDO i Apples AppStore og i Google Play ligesom de øvrige apper de er vant til at finde og bruge dér.

I den indledende fase analyserede vi krav og behov ud fra eksisterende undersøgelser og vores egne antagelser om brugerne og de situationer appen skal fungere i. Det vi især hæftede os ved, var at mobilbrugere der bruger onlinetjenester og apper, tilhører det yngre segment (det fremgår bl.a. af den statistik som Danmarks Statistik udsender hvert år om it-anvendelse i befolkningen¹) med dertilhørende karakteristisk adfærd: De bruger i højere grad end gennemsnittet også sociale medier, spil, underholdning osv. Det var noget vi burde indtænke i appudviklingen. En anden antagelse var at mobiltelefonen oftere bruges udenfor arbejdssammenhænge og derfor kan indebære et andet informationsbehov: Der er i højere grad brug for hurtig, klar og overskuelig besked: Hvordan staver man til x, hvad betyder y, hvad

¹ <www.dst.dk/da/Statistik/Publikationer/VisPub.aspx?cid=018686>.

kommer z af? Når samtidig de mobile enheder har en mindre skærm, taler det for færre, men centrale oplysninger.

Enhver løsning af opgaver som denne er resultatet af en afvejning mellem hensynet til økonomien på den ene side og de tekniske muligheder på den anden. I dette tilfælde var de økonomiske midler stærkt begrænsede da udviklingen skulle foretages indenfor den ordinære bevillingsramme. Det var derfor nødvendigt at fokusere på det absolut nødvendige og på at tænke i økonomiske løsninger: Enkel vedligeholdelse måtte prioriteres over komplicerede løsninger, fx når det gælder opdateringer, og eksisterende web-services der allerede var i anvendelse på hjemmesideversionen, skulle genbruges med så få tilpasninger som muligt. Dette har til gengæld gjort det nemt at bruge DDO-appen som skabelon for selvstændige apper af DSL's øvrige ordbøger.

Et andet valg gik på om appen skulle kunne bruges både online og offline. Vi valgte onlineløsningen selvom nogle brugere givetvis er utilfredse med det valg. Det betyder jo at den eksempelvis ikke fungerer i sommerhuset og hvor man ellers færdes uden netadgang, eller i eksamenssituationer hvor en skole i henhold til gældende regler har spærret for netadgangen. Selvom det teknisk er muligt at lave en fungerende offlineløsning, blev den alligevel fravalgt fordi den ville være mere kompliceret og dermed dyrere at udvikle. Med en online app kan al søgefunktionalitet udvikles ét sted (på vores server) og benyttes på forskellige platforme. Af samme grund ville det være vanskeligere at opdatere en offline app med nye ord: Vi ville blive nødt til at udgive en ny version af hver app frem for – som nu – blot at opdatere på vores server.

Endelig måtte der træffes et valg mellem hvilke operativsystemer appen skulle fungere under. I den ideelle verden vil man gerne servicere alle brugere uanset hvilken enhed de bruger, men samtidig er det også et vilkår at der skal udvikles forfra hver gang et nyt styresystem skal understøttes. Derfor er det nødvendigt at forholde sig til fordelingen af mobile styresystemer.

I Danmark havde Android en markedsandel på 63 % i 2013, mens iOS tegnede sig for 33 %. De resterende 4 % var fordelt på flere mindre operatører, hvoraf Windows var størst med 3 %². Med de begrænsede økonomiske

² Kilde: <www.dr.dk/nyheder/viden/tech/danskerne-foretraekker-google-mobiler>.

midler der var til rådighed, er det ikke svært at indse at Android og iOS var de styresystemer det var nødvendigt at prioritere. Hvis man tilhører de resterende 4 % af mobilbrugerne der har Windows Phone, Symbian eller et andet system, er det selvfølgelig skuffende, men det må ikke opfattes som udtryk for en kvalitativ præference fra vores side, alene som et nødvendigt valg på baggrund af systemernes udbredelse.

3. Appens funktioner

Der er to klare forskelle mellem en hjemmeside og en app: Udover størrelsen på skærmen, som allerede er nævnt, er der den måde brugerne navigerer på, dvs. med fingrene på en berøringsfølsom skærm i stedet for med tastatur og mus. Begge dele stiller nye krav til hvordan et opslag i en ordbog skal udtænkes.

3.1. Et spørgsmål om at skære fra

Først og fremmest handler det om at skære fra. Det første vi tog stilling til, var antallet af ordbøger i den samme app. Vores hjemmeside, ordnet.dk, samler to ordbøger, DDO og Ordbog over det danske Sprog (ODS) og et korpussøgeværktøj, KorpusDK, men vi mente at det ville blive lettere at designe en simpel app hvis den kun indeholdt én ordbog³. Flere ordbøger ville både betyde mere navigation inde i appen og måske risiko for at tabe overblikket over præcis hvilken ordbog man slår op i.

Indholdet i ordbogsartiklerne måtte også skæres til. DDO indeholder bl.a. oplysninger om ords stavning, bøjning og betydning – men også mange oplysningstyper der kan antages at være rettet mod specialister, fx valensoplysninger og grammatiske sprogbrugsoplysninger som ”kun i passiv og i præteritum participium brugt som adjektiv”. Stine Johannsen udførte i efteråret 2011 en brugerundersøgelse af DDO på ordnet.dk (Johannsen 2011,

³ Udover de to ordbøger på ordnet.dk er foreløbig tre andre historiske onlineordbøger offentliggjort på samme platform: Holbergordbog, Moths Ordbog og Meyers Fremmedordbog. ODS og de tre andre historiske ordbøger er senere udkommet som selvstændige apper, men er linket til DDO og hinanden i et netværk, se Troelsgård & Sørensen (2015).

Lorentzen & Theilgaard 2012), og en af konklusionerne var at det som brugerne især efterspørger i ordbogen, er information om ords betydning (herunder synonymer), stavning, bøjning, udtale og etymologi. Desuden nævner brugere at de anvender citater når de ikke umiddelbart forstår definitionerne. Der var derfor disse oplysninger vi prioriterede da vi designede visningen af DDO til appen. Det betød en forsimplet ordbogsindgang, men med de mest centrale oplysningstyper bevaret.

De oplysningstyper vi fjernede, var: orddannelser (sammensætninger og afledninger), valens og øvrige grammatiske oplysninger, sprogbrugs-eksempler (kollokationer) samt kilder for citaterne. Mht. udtale fjernede vi lydskriften, men tilføjede efter et par måneder muligheden for at få opslagsordet læst op. På den måde mener vi at have skabt et bedre overblik over en ordbogsartikel på en lille skærm uden at miste de væsentligste oplysninger.

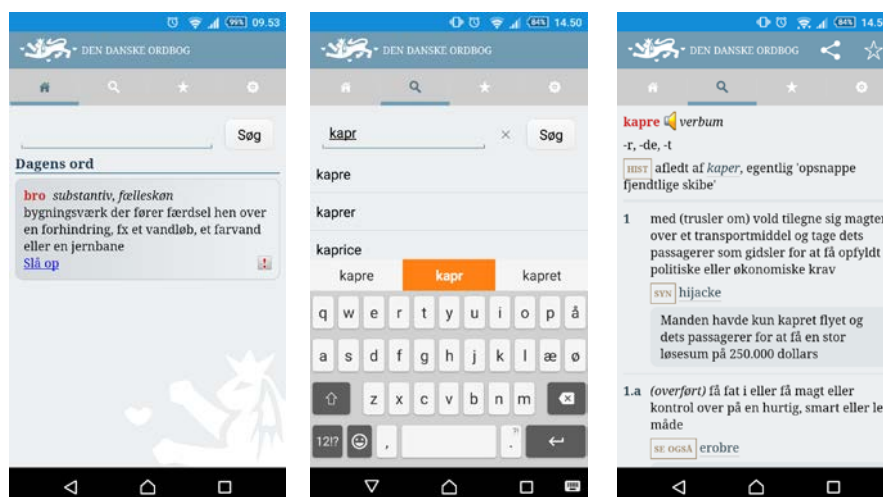
3.2. Den simpleste vej til en søgning – appens opbygning⁴

På en lille skærm har man mindre overblik, og samtidig tager det tid at taste på et virtuelt tastatur i forhold til et fysisk. Man har ikke fordel af at kunne overskue mange informationer samtidig på skærmen og ikke meget plads til at dedikere områder til særskilte oplysningstyper, fx navigation eller søgefelt.

I en app er der forskellige måder at skabe overblik og struktur i indholdsnavigationen på – typisk sker det ved hjælp af såkaldte faneblade (placeret over indholdet i Android og under på iOS), som er anvendt i vores app. Det første faneblad er forsiden, der er holdt bevidst simpel med udelukkende et søgefelt og Dagens ord, som hver dag bliver opdateret med et aktuelt ord eller udtryk. Ved aktivering af søgefeltet bliver man ført til næste faneblad, hvor autofuldførelseslisten automatisk foreslår brugeren ord der begynder med de indtastede tegn, og hvor ordbogsopslaget vises (jf. Figur 1). Vi har valgt ikke at bruge kostbar skærmlads på fx at vise søgefeltet i toppen eller bunden af skærmen, hvilket ses i mange andre ordbogs-

⁴ Appens opbygning er udviklet af Sylvain Fay-Châtelard, Dviance, mens søgeresultatet er udviklet af DSL.

apper⁵, men brugeren kan altid vende tilbage til søgningen eller søgefeltet vha. fanebladene.



Figur 1: Den simpleste vej til en søgning.

Den enkle tilgang til søgning bevirker til gengæld at væsentlige elementer fra hjemmesidens søgeresultat har måttet udelades: dels en liste hvor man kan se andre ord i den alfabetiske nærhed af opslagsordet, dels muligheden for at navigere i et søgeresultat ved siden af visningen af en ordbogsartikel. Begge disse lister er af pladshensyn sparet væk i appen. Det betyder at muligheden for søgning med jokertegn desuden måtte fjernes fra appen, og at der måtte foretages en grundlæggende ændring i tilfælde af homografi (se afsnit 3.3).

De to sidste faneblade i appen er Favoritter og Indstillinger. Favoritter er siden hvor man kan se de opslag som brugeren har valgt at gemme. Dette er en standardfunktion i mange apper som vi valgte at implementere fordi vi regnede med at brugerne ville have en forventning om at finde den. Det samme gælder muligheden for at dele et ordbogsopslag på fx Facebook og Twitter. Vi har endnu ikke undersøgt i hvor høj grad disse standard-

⁵ DDO-appen bruger til gengæld pladsen i toppen af appen til at vise logo og ordbogens navn. Det kan selvfølgelig diskuteres om dette er en vigtigere funktion for brugeren end endnu nemmere adgang til fx et søgefelt.

funktioner bliver anvendt, og de har vist sig at være relativt resursekrævende at udvikle og fejlrette. Hvis vi skulle starte forfra med udviklingen med den viden vi har nu, ville vi derfor overveje at udelade dem og i endnu højere grad koncentrere os om den simplest mulige søgning og den simplest mulige app.

3.3. Homografer

På hjemmesiden kan vi vise en homografartikel og samtidig benytte søgeresultatet til højre for ordbogsartiklen til at vise brugeren at der findes andre ord med samme stavning som kan vælges.

Uden et særskilt søgeresultat måtte vi omtænke en sådan homograf-søgning. Løsningen er blevet at vi viser alle homografer der matcher søgningen, under hinanden, hvilket betyder at et ordbogsopslag i appversionen af DDO ikke længere viser præcis ét lemma. Dette er en principiel ændring fra måden vi betragter et ordbogsopslag på hjemmesiden – og som det betragtes i mange andre onlineordbøger, bl.a. på Gyldendals ordbogssite og Macmillan English Dictionary Online. Opslaget med flere lemmaer svarer til gengæld til hvad fx Ordbogen.com gør.

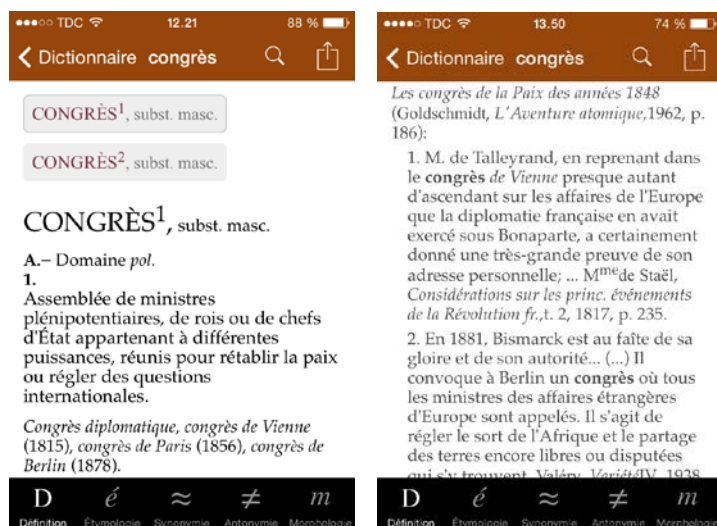
For at signalere at der er flere lemmaer i opslaget, har vi tilføjet en oversigt over de matchende lemmaer først i visningen fordi det ellers kun vil være det første lemma der er synligt. Dette kan desuden bruges til at navigere direkte til et opslagsord og kan dermed siges at svare til søgeresultatet i højre kolonne på hjemmesiden – forskellen er at man kan vælge at scrolle igennem alle lemmaerne på én gang. Dette mener vi fungerer så godt at vi har planer om at indføre noget tilsvarende på hjemmesiden, hvor det netop har voldt mange brugere problemer dels at vælge den rigtige homograf i højre spalte, dels overhovedet at erkende at der eksisterer flere ord der matcher søgningen.

4. Sammenligning med andre ordbogsapper

I dette afsnit ser vi lidt nærmere på to andre, udenlandske ordbogsapper der kan minde om vores DDO-app. Trésor de la Langue Française (TLF) er ligesom ODS en stor nationalordbog og dækker det franske sprog fra 1789

til ca. 1960. Den udkom i 16 trykte bind fra 1971 til 1994 og har desuden været tilgængelig i en onlineudgave, TLFi, i lidt mere end 10 år. Appen henter og præsenterer data fra denne onlineversion.

Når man aktiverer appen, får man vist en lidt kedelig og nøgtern forside med overskriften Dictionnaire og et søgefelt. Herunder findes en kort tekst om at ordbogen er baseret på TLFi, og i et gråt felt ser man logoer for tre involverede institutioner. Det er ikke muligt at finde oplysninger om udvikleren eller mere tekniske oplysninger.



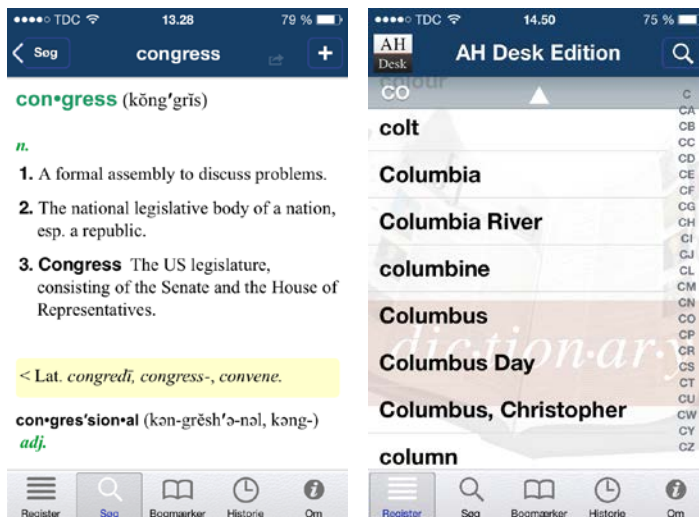
Figur 2: Opslaget *congrès* fra TLF-appen.

I Figur 2 ser vi et opslag, ordet *congrès*. Som det fremgår, er der tale om en ret kompakt visning af den lange definition og de lange belæg. Der er dog gjort et forsøg på at differentiere præsentationen af oplysningerne, fx fremtræder opslagsord og fagmarkering tydeligt (Domaine *pol.*), og det samme gælder kildeangivelsen til citaterne. Det er en stor fordel med fanebladene i bunden af skærmen, idet de giver mulighed for hurtigt at få adgang til særlige oplysningstyper som etymologi eller morfologi. Disse vises så særskilt. En særlig interessant type er synonymi. Her ser man en række synonyme til opslagsordet, og de er ordnet således at man øverst får præsenteret de mest relevante synonyme sammen med en grafisk angivelse af deres grad af synonymi med opslagsordet.

Den anden app vi vil nævne, er American Heritage (AH). Den minder i omfang mere om DDO, Svensk ordbok og til dels Nynorskordboka og Bokmålsordboka. Appen bygger på fjerde udgave af ordbogen. Når vi sammenligner med TLF og DDO, er det slående at vi her møder en mere indbydende og informativ forside, der bl.a. oplyser at ordbogen indeholder over 70.000 opslagsord, mere end 1.000 nye ord og betydninger og ekspertudsagn om korrekt sprogbrug.

I modsætning til de to andre apper er der her adgang til omfattende omtekster, fx en kort beskrivelse (Description) og nøgle til udtale-angivelserne (Pronunciation Key). Bag overskriften iOS Application Help gemmer sig yderligere en del hjælpetekster, bl.a. om bogmærker og avanceret søgefunktion.

Et opslag kan se ud som i Figur 3 (til venstre), luftigt og kortfattet. Der gives ingen belæg, men til gengæld er der lydskrift. Også her finder vi faneblade i bunden. Under overskriften Register finder man en alfabetisk liste over ordbogens opslagsord (Figur 3, til højre), og her fremgår det at ordbogen, som det ofte er tilfældet med engelsksprogede ordbøger, også indeholder proprier som *Columbus* og *Columbia*.



Figur 3: Opslaget *congress* og alfabetisk liste i AH.

5. Modtagelse

DDO-appen er generelt blevet mødt med positive reaktioner og vandt kategorien News & Reference ved Danish App Awards 2013. Vurderingerne i AppStore og Google Play ligger klart i den høje ende: I AppStore får den 4½ stjerne af 5 mulige, og i Google Play får den 4 ud af 5.

Blandt brugerkommentarerne er der nogle rosende bemærkninger som vi selvfølgelig sætter pris på. For eksempel er man glad for at appen er gratis, man mener at den fungerer som den skal, og det er godt at man kan have ordbogen med sig. Funktionen Dagens ord giver anledning til positive kommentarer, men også til nogle mere kritiske: Nogle vil gerne have mere pudsige og sjældne ord i stedet for ord de godt kender i forvejen, og andre vil også have Dagens ord vist i den daglige notifikation i stedet for bare at blive mindet om at det findes. Foreløbig har vi af resurse-mæssige hensyn nemlig ikke prioriteret at sende brugerne Dagens ord som en egentlig pushbesked selvom det teknisk er muligt. Andre reaktioner går på at man gerne vil have lydskrift ligesom på hjemmesiden.

En del brugere efterlyser også en offline app, bl.a. for at kunne bruge den til eksamen hvor man ikke nødvendigvis har adgang til internettet. Vi er selvfølgelig klar over denne problemstilling, men har som nævnt i afsnit 2 valgt at lade appen hente data fra en webservice. Og når det gælder brug til eksamen, er det et område der er under hastig forandring i disse år. Om et par år er det måske slet ikke nødvendigt at diskutere det spørgsmål mere.

Vores ordbogsapper, især DDO, er også blevet genstand for en grundig granskning i LexicoNordica 21 ved Holmer & Sköldbberg (2014). De er også generelt positive overfor DDO-appen, bl.a. fremhæver de på plussiden at de grundlæggende oplysninger er med, at der er mulighed for kommunikation mellem brugeren og udvikleren, og at appen linker til netversionen af ordbogen. De nævner dog som et problem at hvis man følger linket, kommer man ikke til en mobiltilpasset visning (som det kendes fra Wikipedia og mange andre hjemmesider). Vi vil meget gerne have en mobilvisning og håber på at der bliver økonomiske muligheder for det i fremtiden.

Holmer & Sköldbberg har andre ønsker til DDO-appen, og det er punkter der også står på vores ønskeliste. For eksempel vil de gerne have mere omtekst, og det synes vi også ville være godt. Her kan American Heritage-appen måske tjene til inspiration. Engelsksprogede apper er ligeledes

længere fremme når det gælder et andet ønske, nemlig stemmestyring så man kan tale til appen. Endelig ønsker Holmer & Sköldberg sig også at flere ord i artiklerne bliver hyperlinks. Hertil kan man sige at synonymmer, antonymer og andre beslægtede ord allerede nu i høj grad er links, og på nyere iPhones har vi faktisk allerede gjort det muligt at slå et hvilket som helst ord i ordbogsartiklen op alene ved at ”lang-klikke” på det.

Den Danske Ordbogs app er siden lanceringen i 2013 installeret i alt 119.000 gange på iOS-enheder og 58.000 gange på Android-enheder⁶. Det lavere tal for Android-plattformen, på trods af styresystemets større udbredelse, er noget vi skal undersøge nærmere: Skyldes den udelukkende en generel forskel i brugernes adfærdsmønstre på de to platforme, eller kan vi gøre noget for at trænge bedre igennem? Samtidig vil vi gerne fortsætte den gode gennemslagskraft i iOS-verdenen, ligesom vi holder øje med andre styresystemers udbredelse i Danmark.

Litteratur

Ordbøger

AH = *American Heritage® Desk Dictionary* (app), Fourth Edition (2003, 2001, 2008). Published by Houghton Mifflin Harcourt Publishing Company.

Bokmålsordboka. Browserversion: <bokmalsordboka.uio.no/>

DDO = *Den Danske Ordbog* (app). Browserversion: <ordnet.dk/ddo>

Gyldendals ordbogssite. Browserversion <ordbog.gyldendal.dk>

Macmillan English Dictionary Online. Browserversion <macmillandictionary.com>

Nynorskordboka. Browserversion: <www.nob-ordbok.uio.no>

ODS = *Ordbog over det danske Sprog*. Browserversion: <ordnet.dk/ods>

Ordbogen.com. Browserversion: <ordbogen.com>

Svensk ordbok utgiven af Svenska Akademien (2009). Norstedts i distribution.

⁶ Tal fra september 2015. Af de 58.000 installationer på Android er 36.000 ”nuværende”. Et tilsvarende tal findes ikke for iOS.

TLFi = *Trésor de la Langue Française informatisé* (app). Browserversion:
<zeus.atilf.fr/tlf.htm>

Anden litteratur

Holmer, Louise & Emma Sköldbberg (2014): Appifiering til allas lycka? Om danska ordboksappar med särskilt fokus på DDO. I: *LexicoNordica* 21, 235-252.

Johannsen, Stine (2011): *Brugerundersøgelse af Den Danske Ordbog for DSL*. Upubliceret rapport, Det Danske Sprog- og Litteraturselskab.

Lorentzen, Henrik & Liisa Theilgaard (2012): Online dictionaries – how do users find them and what do they do once they have? I: Ruth Vatvedt Fjeld & Julie Matilde Torjusen (eds.): *Proceedings of the 15th EURALEX International Congress*, Department of Linguistics and Scandinavian Studies, University of Oslo, 654-660.

Troelsgård, Thomas & Nicolai Hartvig Sørensen (2015): Nyt liv til støvede ordbøger – historiske ordbøger på nye platforme. I: Asgerd Gudiksen & Henrik Hovmark (red.): *Nordiske Studier i Leksikografi 13* (Skrifter udgivet af Nordisk Forening for Leksikografi 14) (dette bind).

Henrik Lorentzen
seniorredaktør

Det Danske Sprog- og Litteraturselskab
Christians Brygge 1
DK-1219 København K
hl@dsl.dk

Nicolai Hartvig Sørensen
seniorredaktør

Det Danske Sprog- og Litteraturselskab
Christians Brygge 1
DK-1219 København K
nhs@dsl.dk

Lars Trap-Jensen
ledende redaktør

Det Danske Sprog- og Litteraturselskab
Christians Brygge 1
DK-1219 København K
ltj@dsl.dk

Gamle ordbøker og digitale utgaver

Christian-Emil Smith Ore

Today it is relatively easy to re-publish old dictionaries on the web based on retro digitized text. To create trustworthy editions which can serve as the fundament for reliable Linked Data, this has to be done with care and it is necessary to follow the basic principles of digital scholarly text editions. A digital edition of the Norwegian Dictionary written by the Norwegian linguist Ivar Aasen published in 1873 is used as example.

1. Innledning

I denne artikkelen vil jeg diskutere en del forhold som jeg mener er av vesentlig betydning for digital utgivelse av gamle ordbøker. Utgangspunktet er det arbeidet som Oddrun Grønvik og jeg har gjort for å forberede en felles digital utgave av ordbøkene til de to norske språkforskerne Ivar Aasen og Hans Ross. Det er spesielt fire forhold jeg vil behandle:

En eldre ordbok er ikke en leksikalsk database i forkledning. Det gjelder i høy grad Aasens ordbøker. De er pionerordbøker med mye forklarende tekst.

Gamle ordbøker har en klar egenverdi som dokumentasjon av språket og kulturen på den tiden de ble skrevet, og er dermed gode verktøy til å forstå tekster fra denne tiden. Når eldre ordbøker utgis på nytt, må den originale teksten derfor presenteres fullstendig og uendret. Dersom en likevel ønsker å legge til tekst eller flytte på eksisterende tekst, må det være tydelig markert.

Papir og skjerm er to forskjellige medier. For en digital utgave er det hensiktsmessig å redigere ordboksteksten i mindre, selvstendige deler som kan leses uten kontekst. Eldre ordbøker har ofte et preg av å være skrevet som en sammenhengende tekst. For eksempel vil mindre ordartiklene for avledninger og sammensetninger ofte forutsette at en lett kan se hovedartikkelen for grunnordet. Det er verken god formidling eller innenfor god edisjonsfilologisk praksis å foreta en finoppdeling som bryter med ordboksverkets egenart.

De edisjonsfilologiske kravene og den eldre tekstens spesielle krav til visning er ikke til hinder for en dypere analyse og tilhørende oppmerking av

informasjonselementer som for eksempel normaliserte oppslagsord, ordklasse, stedsnavn, kryssreferanser og henvisning til andre verk. En slik oppmerking bør være obligatorisk siden den gir muligheter til å sammenstille opplysningene i mange ordbøker. Men analysen og oppmerkingen er en del av det moderne edisjonsfilologiske kommentarapparatet og ikke en del av det originale verket.

2. Aasens ordbok fra 1873 – posisjon, innhold og kategorier

I *Norsk Ordbog med Dansk Forklaring* som utkom i 1873 (Aasen 1873), legger den norske språkforskeren Ivar Aasen (1813–96) frem et utkast til et standardisert norsk skriftspråk basert på det samtidige talemålet. Ordboken bygger på Aasens omfattende innsamlingsarbeid i Norge. Aasens ordbok fra 1873 er et grunnleggende verk i studiet av norsk språk. Den bygger på hans *Ordbog over det norske Folkesprog* (1850) og på *Norsk Grammatik* (1864).

Filologen Hans Ross (1833–1914) fulgte opp Aasens innsamlingsarbeid og publiserte i 1895 ordboken *Norsk Ordbog, Tillæg til "Norsk ordbog" af Ivar Aasen*. Ross utgav seks supplement til sin egen ordbok i perioden 1895 til 1913. Ross' ordbok er redigert som et supplement til Aasens ordbok fra 1873 og må leses i sammenheng med denne.

Aasens ordbøker men også Ross' ordbok med tilleggene er svært viktige kilder for arbeidet med *Norsk Ordbok* (NO). De er ikke ekserpert i seddelarkivet til *Norsk Ordbok*, men de er tilgjengelige gjennom det såkalte grunnmanuskriptet (GMANUS) som finnes som en del av NO2014-prosjektets omfattende bakgrunnsmateriale. Grunnmanuskriptet er et maskinskrevet utkast til *Norsk Ordbok* fra 1935. En vesentlig del av dette manuskriptet utgjøres av ordbøkene til Aasen og Ross, der ordboksartiklene med de danske forklaringene er oversatt til norsk, tilpasset og flettet sammen til nye felles ordboksartikler. Grunnmanuskriptet inneholder altså en oversettelse og bearbeidelse av de originale tekstene til Aasen og Ross.

Det har lenge vært et ønske om å lage en elektronisk versjon av ordbøkene til Aasen og Ross. Men siden de har vært tilgjengelige gjennom Grunnmanuskriptet, de trykte utgavene og som digital brødtekst (se nedenfor), har en slik digital utgave ikke hatt høyeste prioritet i NO2014-prosjektet. Men det har like fullt vært arbeidet med å gjøre dem tilgjengelig i

databasene for NO2014-prosjektet. Tidlig i 2014 foreslo Oddrun Grønvik og jeg å lage et prosjekt for en digital utgave av Aasen og Ross' ordbøkene med vedlegg. Vi fikk klarsignal fra ledelsen i NO2014-prosjektet i 2014 og lagde en foreløpig versjon som var ferdig til mai 2015.

Listen nedenfor angir målene til hovedprosjektet vårt. For NO2014-prosjektet var punkt 1 nedenfor det viktigste. Det lenker ordbøkene inn i den felles kildedatabasen. Prosjektets mål er altså å

1. identifisere oppslagsord og lenke dem til moderne form gjennom Metaordboka, indekseringssystemet for språksamlingene
2. identifisere stedsnavn og tolke dem som en referanse inn i ett eller flere av herredene i 1947 og dermed georeferere referansene
3. finne sitater, målføreformer, og interne og eksterne referanser
4. gjøre alle ordbøkene tilgjengelige på nett og som Linked Open Data
5. gjøre oppslagsord, artikkeltekst og geografiske kilder søkbare hver for seg og der en kan bruke moderne rettskriving som søkeargument
6. gi alle tekstene en TEI P5 (TEI 2015) mark-up i tråd med moderne edisjonsfilologiske prinsipper

Det er viktig å tenke gjennom de faglig-prinsipielle sidene ved et slikt prosjekt før og under arbeidet. Oddrun Grønvik har diskutert forhold rundt normeringsproblemene i Grønvik (2015). I denne artikkelen vil jeg fokusere på edisjonsfilologi, bruk av gamle ordbokstekster og formidling i en digital verden. Eksempelene er hentet fra Aasen (1873).

3. Det digitale tekstgrunnet

Da vi begynte prosjektet, fantes grunnteksten i Aasens ordbøker (1850 og 1873) som elektronisk brødtekst. I forbindelse med Ivar Aasen-jubileet i 1996 var tekstene blitt tastet inn og korrekturlest ved Språksamlingene, Institutt for nordistikk og litteraturvitenskap ved Universitetet i Oslo.

Disse tekstene ble siden brukt som grunnlag for trykte utgaver ved K. Kruken og T. Aarset (blandt andre Aasen 2003). Kruken og Aarset har flettet inn tilleggene i hovedteksten og gjort en lett omredigering som dessverre ikke er markert i teksten.

Hans Ross' ordbok med tilleggene ble også tastet inn og korrekturlest ved Språksamlingene i forbindelse med den storstilte digitalisering av kilde-materialet til Norsk Ordbok på 1990-tallet.

Nettstedet bokhylla.no er lagd av Det norske nasjonalbiblioteket og inneholder i dag faksimiler av om lag 500.000 bøker inkludert verkene til Aasen og Ross. Tekstene er OCR-lest, men resultatet er ikke korrekturlest. For bøker satt i fraktur som er tilfellet for Aasen og Ross' verker, er resultatet svært dårlig.

De elektroniske tekstene fra 1990-tallet er basert på førsteutgavene og er komplette, men er typografisk forenklet. I den inntastete teksten er fraktur erstattet med antikva, liten rettskrift med kursiv og uthevet fraktur med halvfet. Antikva i originalen er ikke markert i avskriften. De er det beste utgangspunktet for vårt edisjonsfilologisk orienterte prosjekt, men vi har likevel valgt å legge inn pekere til faksimilene på bokhylla.no slik at det er mulig å konsultere det originale skriftbildet.

4. Leksikalsk database?

For 20–35 år siden hadde man, spesielt innenfor datalingvistiske miljøer, en viss forhåpning om at en med utgangspunkt i eksisterende ordbøker kunne lage leksikalske databaser for datalingvistiske anvendelser. Den grunnleggende tanken var at den grammatikalske informasjonen og ikke minst betydningsinndelingene i ordbøkene representerte en objektiv beskrivelse av det aktuelle språket, (Kilgariff 1993). Dette var ikke en urimelig antakelse, siden mange ordbøker har vært og fortsatt betraktes som en autoritativ beskrivelse av et språk, fordi ordbøker er ordnet i artikkelform med overskrift og en del faste kategorier og likner på en databaseutskrift, og ikke minst fordi en ordbok representerer en svært stor investering i form av arbeidstid.

Men som leksikografen Sue Atkins har bemerket¹, ville det være en bedre ressursanvendelse om en ga en gruppe leksikografer i oppgave å lage en leksikalsk database fra bunnen av enn å la en gruppe datalingvister prøve å omforme en ordbok til en leksikalsk database. Ingen pre-digital ordbok er systematisk nok eller har klare nok grenser mellom tekstkategoriene til å være egnet til det. Gode og forståelige beskrivelser av ords betydning

¹ I lunsjsamtale under *CompuLex'92*, Budapest 1992.

beregnet på å leses av mennesker, vil sjelden være analytiske nok, jf. Hanks (1979).

5. Dokumentasjon av kulturarv og etterrettelighet

Språk og navn er ikke-materielle fenomener og hører til vår immaterielle kultur(arv) slik som Det norske kulturdepartementet også påpeker:

UNESCO-konvensjonen om den immaterielle kulturarven frå 2003 har mellom anna til formål å verna språk som uttrykksmiddel for denne arven, til dømes språk i form av munnlege tradisjonar og uttrykk. Og i konvensjonen om å verna og fremja eit mangfald av kulturuttrykk [...] er det i dei innleiande merknadene streka under at språkleg mangfald er ein vesentleg del av det kulturelle mangfaldet (St.m.35 2007).

Noen vil kunne innvende at det er farlig å sette språk i bås med kulturarv fordi omverdenen lett vil kunne se leksikografi som en støvete museal og ikke-vitenskapelig aktivitet. Det er selvfølgelig tøv. For det første er samtidskulturen av kort varighet og blir i morgen kulturarv fra i går. Enhver beskrivelse av et språk må være retrospektiv. For det andre er vitenskapelighet knyttet til metode- og etterprøvnbarhet av resultatene. Men også i en vitenskapelig leksikalsk beskrivelse av et språk vil tolkningen av ordenes mening være preget av den kulturelle konteksten forskeren lever i. Ordbøker blir med tiden en historisk dokumentasjon av de kulturelle syn som hersket på tiden de ble skrevet, jf. definisjonen av hottentott som ”innfødt tilhørende en lavtstående syd-afrikansk folkerase” i Norsk Riksmålsordbok (NRO). Eldre ordbøker kan også gi detaljert informasjon om foreldete praktiske forhold som ikke kommer med i ”samtdsordbøker”.

Dette bringer oss inn på et annet aspekt, nemlig ombruk av gamle ordbøker. I forlagsbransjen har det ut fra rasjonalitetsbetraktninger ikke vært uvanlig å foreta en enkel revisjon av eldre ordbøker og presentere dem som nye og oppdaterte. Dette er en praksis som blir mer og mer diskutabel desto lengre tid det har gått siden den originale ordboken utkom. En eldre ordbok er en tekst som i sin tid ble skrevet for å bli lest og brukt av et bestemt

publikum. Å revidere og gi ut en gammel ordbok uten tydelig å klargjøre hva som er nytt og gammel er ikke etterrettelig.

Ordboksredaktørene på 1800-tallet organiserte ikke sine manuskripter etter databasemodeller eller XML-skjema. Men mange eldre ordbøker er skrevet med en imponerende konsistent artikkelstruktur. Både Ivar Aasen og Johan Fritzner (Fritzner 1883) har en veldig ryddig artikkelstruktur i sine ordbøker. Hos Hans Ross er strukturen i artiklene temmelig fri og mye løsere enn hos Aasen. Men i Aasens ordbok fra 1873 ser en også eksempler på at den er skrevet for å leses som en sammenhengende tekst. Se for eksempel 'and' (metafysisk begrep) og 'and' (fuglen) i Aasen (1873). En moderne ordboksdatabase svarer til å klippe opp ordboksteksten til en bunke løse papirlapper. På 1800-tallet falt det neppe noen inn å publisere en ordbok på den måten. Aasen så teksten som en helhet.

6. Ivar Aasens ordbok

Hvor har vi tenkt å plassere arbeidet med Aasens ordbok i dette bildet? Det er et mål for oss å følge moderne edisjonsfilologiske prinsipper. En kan grovt dele edisjonsfilologien inn i den tradisjonelle, der målet er å lage en beste tekst på grunnlag av tilgjengelige tekstvitner, og nyfilologien (new philology, se f.eks. Nichols (1990)) der en ser hvert tekstvitne (manuskript) som et studieobjekt i seg selv. Vårt prosjekt heller i nyfilologisk retning: I den transkriberte teksten skilles det riktignok ikke mellom alle typesnittene i originalen. Linjeskift og spalteskift er ikke markert. Det er derfor ikke mulig å vise den elektroniske teksten i en "faksimilær" gjengivelse. Men dette kompenseres med tilgang til faksimiler av originalen. I den foreløpige digitale utgaven av Aasens ordbok fra 1873 er det bare hovedteksten som gjengis. Aasens egne tillegg og tilleggene til Hans Ross er ikke med. I den fullstendige digitale utgaven vil alle ordbøkene bli utgitt separat i den nyfilologiske tradisjonen.

En kan fundere på om Ross noen gang forestilte seg muligheten av å utgi en samredigert utgave av Aasens ordbok og hans egne tillegg. Det er mulig at han hadde for stor respekt for Aasen til å våge det, at Aasen motsatte seg det eller at det ikke fantes ressurser. I en videreutviklet versjon vil vi eks-

perimentere med en felles utgave basert på hyperlenker og eventuelt sam-sortert visning av de ulike ordbøkene med tilleggene.

7. Tolkning av teksten og indeksering av ord og navn

Som nevnt i innledningen, er formålet å bevare Aasens tekst mest mulig ufortolket og la vår egen tolkning representeres gjennom ekstra TEI-koding av teksten. Et annet mål er å trekke ut informasjon om hva vi mener Aasen anså som gangbare ordformer i ordinventaret i norsk og, om mulig, hvor i Norge Aasen mente de ble brukt. Ordformene eller oppslagsordene vil deretter bli koplet til resten av dokumentasjonsdatabasen. Denne koplingen vil gå gjennom den felles ordindeksen, Metaordboka, og dermed også til selve *Norsk Ordbok*.

skorra, v.n. (ar), glide paa Iis (= skreida, renna seg). Smaal. Nt. schurren. (Schambach 188). Ved Mandal **sko-renna**: glide med bare Sko, uden Skøiter.

```
<entry type="hel" n="26232 1" xml:id="A2_orig27065"><form
n="A2_orig33271" type="origlemma">skorra</orth>,
<gramGrp><pos>v.n.</pos></gramGrp> <form n="A2_orig33272"
type="inflected_"> <orth>(ar)</orth></form></form>, <sense><def> glide paa
Iis (= skreida, renna seg). <placeName xml:id="p_026157" n="5,
Austf">Smaal.</placeName> <rs type="lang">Nt.</rs> schurren.
(<ref><bibl><title>Schambach</title> 188</bibl></ref>). Ved <placeName
xml:id="p_026158" n="323, A-VAgd">Mandal</placeName> <hi
rend="bold">sko-renna</hi>: glide med bare Sko, uden Skøiter.
</def></sense></entry>
```

Figur 1: Artikkelen for oppslagsordet *skorra* vist som tekst og vist med full xml-oppmerking.

En slik kopling vil kunne gi leseren et inntrykk av at Aasen regnet ordformen som en del av hans utkast til det norske skriftspråket. Men egentlig er det vår tolkning av teksten som ligger til grunn. Å finne årsaken til at et stedsnavn er referert i en artikkel, er adskillig mer problematisk. Angir stedsnavnet maksimalutbredelse av et oppslagsord eller at Aasen har

observert en bestemt dialektrealisering av ordet der? Det er viktig å ha en gjennomtenkt forståelse av hva som skal tagges.

Figur 1 viser et eksempel på tagging av en ordartikkel. Vi kan regne med at hovedoppslagsord står på sin vanlige plass først i en artikkel. Videre kan vi regne med at ordformer med ordklassemarkering i gitte omgivelser kan sees som et oppslagsord. Det er viktig siden vi baserer oss på et halv-automatisk utplukk av oppslagsordkandidater. Nærlesing av teksten viser at ikke alle uthevede ordformer er egne oppslagsord, og det finnes ord som ikke er uthevet men som likevel kan regne som oppslagsord. I mange tilfeller kan en bruke signalord og signalfraser til å lete dem opp. Ordet ‘hertil’ indikerer gjerne at neste ord er enten en sammensetning der hovedoppslagsordet i artikkelen inngår som ledd eller at det kommer en avledning av hovedoppslagsordet. I eksempelet i Figur 2 er det derfor rimelig å tolke ordet *tjeldvev* i artikkelen for *tjeld* som et selvstendig oppslagsord. Aasen bruker ‘I Sammensætning’ som innledning til en liten forklaring om hvilken form hovedoppslagsordet har som ledd i sammensetninger for så å ramse opp noen eksempler på slike, se artikkelen for *Kvitel* i Figur 3. Ingen av sammensetningene som nevnes der har egne artikler i ordboken, og de er vel av Aasen betraktet som vanlige, transparente sammensetninger som han har observert. At den første er ført med ordklasse og definisjon, er vel mest for å skille den fra *garn* i betydningen nett.

Tjeld, n. 1) Telt, Hytte af Seildug osv. Nordl. i Formen **Kjeld** (Kjell), især om et Skjul af udspændte Seil. (Andre St. kaldet “Telt”, som er den nedertydske Form). G.N. *tjald*; Sv. tjäll. – 2) et Forhæng, et hængende Tæppe. Hall. (Kjeld). Ogsaa: et Tørklæde, en Dug at kaste over Skuldrene. Rbg. – 3) et Sengetæppe (= Kvitel), Uldtæppe at ligge paa (eller egentl. at indsvøbe sig i). Sæt. Tel. Num. (Kjeld). Hertil **Tjeldvev** (Kjeldvæv), m. Tæppevæv.

Figur 2: Artikkelen for oppslagsordet *tjeld*.

Formen *Kjeld* i Figur 2 er uthevet av Aasen. Det finnes en egen artikkel for *Kjeld* med henvisning til *Tjeld* i betydningen telt eller teppe. Så her er det en kryssreferanse. Formen *sko-renna* i Figur 1 er også uthevet av Aasen, men her finnes det ikke en egen artikkel. Utplukksprogrammet er forsiktig og har ikke markert noen av disse formene som oppslagsformer. Formen *sko-renna*

er eget oppslagsord i *Norsk Ordbok*. Det skyldes vel at det finnes andre belegg for det i seddeldatabasen til NO2014.

I artikkelen for *kvitel* (i Figur 3) er *Gvitel* uthevet av Aasen. Men denne formen er ikke et oppslagsord, men en avvikende målform fra Numedal som Aasen mener det er verd å nevne. I samme artikkel står det 'Ellers kaldet Blæja og Tjeld'. Dette kunne indikere at det er to dialektformer Aasen ikke betrakter som oppslagsordkandidater. Men begge har egne artikler i ordboken. Konklusjonen er at et slikt utplukksarbeid med hell kan gjøres ved hjelp av en utplukksalgoritme, men tvilstilfellene må kontrolleres av en kompetent leksikograf. Oddrun Grønvik (Grønvik 2015) har beskrevet dette i mer detalj.

Kvital (i'), m. Sengeteppe, Uldteppe at ligge paa. B. Stift, Vald. Hall. Tel. og flere. Aftvig. **Gvital**, Num. (Jf. Brotkvitel, Mjukkvitel). G.N. *hvitill*. Eng. Dial. *whittle*. Ellers kaldet Blæja og Tjeld. Fleertal **Kvitlar** (nogle Steder Kvislar). I Sammensætning tildeels Kvitle (Kvitla), som **Kvitlegarn**, n. (Traad til Sengetæpper), **Kvitlevarp**, **Kvitlevest** og flere.

Figur 3: Artikkelen for oppslagsordet *kvitel*.

Rene kryssreferanser er ganske greie å identifisere siden en her har "fasiten" i form av den komplette oppslagsordlisten. En kan også plukke ut referanser til ord i andre språk og til litteratur som Aasen viser til i ordboksteksten.

Det er mange stedsangivelser i Aasens ordbok, om lag 37 000. Men det er langt fra slik at en kan regne med at stedsangivelsene gir noen oversikt over ordenes utbredelse. Alminnelige ord med utbredelse over hele Norge er ikke geografisk markert. Det synes mer å være spesielle betydninger eller spesielle former som er angitt med sted. Stedsangivelsene i Aasen kan ikke brukes til å lage utbredelseskart for enkeltord. Vi har likevel tatt oss bryet med å tagge stedsnavnene. De er så knyttet til herredsinndelingen i Norge i 1947 som er det året antall herreder var på sitt høyeste, noe over 700. Denne geografiske inndelingen av Norge brukes også i *Norsk Ordbok*. Herredene følger topografien og ferdselsårer. Det er sjelden at vesentlige dialektfenomener deler et herred i to. En kan derfor bruke herredene som geografiske byggeklosser. En del av stedsbetegnelsene hos Aasen refererer til fogderier. Et fogderi var delt i herreder som senere kan ha blitt delt i mindre herreder inntil denne inndelingen nådde sitt maksimum i 1947. Ved å bruke

1947-inndelingen kan en med stor nøyaktighet tilordne et eller flere herreder til stedsnavnene i Aasen (1873). I den elektroniske utgaven er det mulig å søke både på originale stedsnavn, herredsnavn fra 1947, fylker og landskap. Som en hjelp for leseren er det også lagt inn "mouseover" slik at det moderne navnet eller en liste av slike dukker opp.

8. Hva skal vi være forsiktige med?

En kan ikke alltid være sikker på hvorfor Aasen har ført opp en dialektform og i hvilken relasjon den står til resten av artikkelteksten. Det er heller ikke alltid mulig å vite hva et stedsnavn indikerer. Det kan være vanskelig å gi artiklene en streng formell struktur slik vi gjerne gjør det i et moderne XML-basert redigeringsystem. Likevel kan en indre analyse være formålstjenlig om man vil vise større artikler på en mer pedagogisk måte eller man vil prøve å flette inn Ross' tillegg i Aasens ordbok. I den grad vi skal prøve å gjøre det, må resultatet av en slik strukturanalyse alltid sees som en subjektiv analyse og markeres som en annotasjon til den originale ordboksteksten.

Det er den originale ordboksartikkelen som er den grunnleggende enheten og som må stå i sentrum. All koding og alt uttrekk til lister over oppslagsord og stedsnavn er hjelpemidler for å få bedre tilgang til informasjonen i ordboken. Dette strenge synpunktet utelukker ikke at vi kan bruke informasjonen i ordboken i et større sammenlenket informasjonssystem som for eksempel databasen (Metaordboka) over kildematerialet i NO2014-prosjektet. Vi må bare være nøye med å oppgi hvilken tolkning av teksten informasjonen er basert på. Dette er normal vitenskapelig kildekritikk. Men i en digital kontekst og spesielt på internett er det ikke alltid så lett å skille mellom originalinformasjon og avledet informasjon. Det er heller ikke noen sterk tradisjon for å være nøye med det på nettet.

9. Hvordan tilrettelegge ordboken til digital utgave?

I en digital utgave av en ordbok, gammel eller ny, er det vanlig at grensesnittet gir brukeren anledning til å søke blant oppslagsord, men noen ganger også foreta fritekstsøk i artikkeltekstene. Man får da gjerne opp en resultatliste med utsnitt av de aktuelle artiklene og eventuelt også en

alfabetisk liste over oppslagsord før og etter den artikkelen som en ønsker å se i sin helhet. Et slikt grensesnitt betrakter en ordbok som en samling av uavhengige og frittstående artikler. Men for eldre ordbøker, som oftest er forfattet som en sammenhengende tekst med kryssreferanser av typen ”se også nedenfor”, bør en ha en mulighet til å rulle forover og bakover i den retro-digitaliserte ordboksteksten og se flere artikler på en gang. Det bør også være mulig å få se faksimiler av den trykte originalen. Det siste er vesentlig for etterretteligheten til digitale utgaver, selv om den digitale utgiveren har brukt mye tid på å gjenskape originalen som digital tekst.

Aasen (1873) har en norsk ortografi som skiller seg vesentlig fra den moderne normen. I tillegg er forklaringen skrevet på et alderdommelig dansk. Stedsnavnene er ikke alltid like forståelig for dagens lesere. Det har derfor vært viktig å kople oppslagsordene til moderne former. Det er gjort gjennom den såkalte Metaordboka som også har manuskriptdatabasen til NO2014-prosjektet og flere moderne ordbøker koplet til seg.

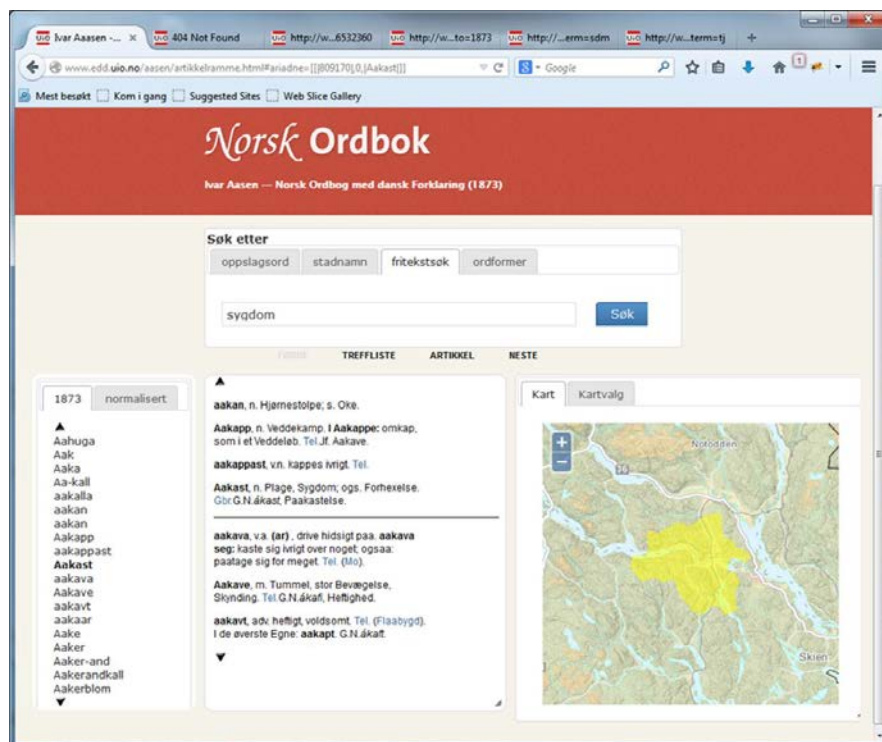
Stedsnavnene er som tidligere forklart, lenket til de norske herredene fra 1947. Nå er ikke det tale om helt moderne enheter, men navnene burde være forståelige. Siden herredsgrensene fra 1947 er digitaliserte og georefererte, kan vi bruke kart både til å vise hvilke(t) område i Norge hvert 1947-herred svarer til og bruke et digitalt kart som søkeinnngang.

I Norge er kunnskapen om 1800-tallsdansk for nedadgående. Som en hjelp kan det derfor være en idé å få tilgang til en fullformsliste over danske ordformer fra 1800-tallet og på den måten muliggjøre enkle oppslag i ODS som kan trigges ved å holde markøren over et ord og for eksempel høyreklikke (eller tilsvarende på en telefon). I ODS er det mange referanser til norske dialektord. Om disse var markert, ville en fra digitale utgaver av Aasens ordbøker eller *Norsk Ordbok* (NO2014) kunne hente dem opp. Tilsvarende kunne en tenke seg at en fra ODS kunne få tilgang til de norske ordbøkene.

I stedet for å lage slike skreddersydde koplingsløsninger for hvert ordbokspar, kunne det være verd å tenke gjennom en litt mer generalisert løsning. De fleste ordbøkene, i alle fall i Norden, inneholder et (varierende) utvalg av de samme informasjonskategorier som oppslagsord, ordklasse, stedsangivelser og litteraturhenvisninger, for å nevne noen. Om en kunne enes om et felles sett av slike kategorier, så kunne en gjennom såkalt ’Linked Data’-grensesnitt åpne for kryss-søk og uthenting av på tvers av

ordbøkene. Det foreløpige grensesnittet for Aasens ordbok er basert på bare tre enkle grensesnitt eller kommandoer for søk etter artikler, stedsnavn eller oppslagsformer. De kan brukes fra andre digitale ordbøker.

- Artikkelsøk: www.edd.uio.no/aasen/a2/artikkel?type=opp&soek=tj*
- Navneliste: www.edd.uio.no/aasen/a2/stadliste?orto=1873&term=s*
- Formliste: www.edd.uio.no/aasen/a2/formliste?orto=1873&term=t*



Figur 4: Enkelt grensesnitt mot den digitale utgaven av Aasens ordbok.

10. Oppsummering

I denne artikkelen har jeg diskutert digital nytgivelse av eldre ordbøker og en del prinsipielle forhold som en må ta i betraktning om digitaliseringen skal yte originalen rettferdigheten og samtidig framstå som både lettbrukt og

etterrettelig. Kort oppsummert: Ved digital nyutgivelse bør en anvende edisjonsfilologiske metoder. En ordbok er et verk som speiler den tiden den ble skrevet i. Endringer i den opprinnelige ordboksteksten representerer nytgiverens subjektive tolkninger og må markeres tydelig. Slike endringer bør sees som annotasjoner og helst legges i teksten som XML-koding eller som 'stand-off' koding i separate filer. En bør tilby faksimiler av originalen. Den digitale utgaven bør ha et åpent grensesnitt slik at andre kan lenke til med såkalte 'Linked Data'-mekanismer. Da kan den digitale ordboken inngå i et større nettverk med de fordeler det gir.

Litteratur

Ordbøker

- Aasen, Ivar (1965): *Norsk Grammatik*. Omarb. Udg. af "Det Norske folkesprogs Grammatik", 3die uforandr. Udg. (Opph. utg. Christiania 1864). Oslo: Universitetsforlaget.
- Aasen, Ivar (1918): *Norsk Ordbog med dansk Forklaring*. Fjerde uforandrede udgave. Kristiania: Vestmannalaget & Cammermeyers forlag.
- Aasen, Ivar (2003): *Norsk Ordbog med dansk Forklaring*. Utg. ved Kristoffer Kruken og Terje Aarset. Oslo: Samlaget
- Fritzner, Johan (1886-1896): *Ordbog over det gamle norske sprog*. Omarbeidet, forøget og forbedret udg. Kristiania: Den norske Forlagsforening.
- GMANUS = *Grunnmanuskriptet* (1935) < www.norskordbok.uio.no >.
- Metaordboka* <www.norskordbok.uio.no>.
- NO = *Norsk Ordbok I-XI* (1950-2013). Oslo: Det norske samlaget.
- NRO = *Norsk Riksmålsordbok I*, annet halvbind (1937). Utarbeidet av Trygve Knudsen og Alf Sommerfelt. Oslo: Aschehoug.
- Ross, Hans (1971). *Norsk ordbog, tillæg til "Norsk ordbog" af Ivar Aasen*. Uforandret opptrykk av utg. 1895 + senere tilleg. Oslo: Universitetsforlaget.
- Ross, Hans (1895-1912): *Tillæg 1-6 til "Norsk ordbog" af Ivar Aasen*. Kristiania: Alb. Cammermeyers Forlag.

Annen litteratur

- Grønvik, Oddrun (2015): Eit lemmaregister for Ivar Aasens Norsk Ordbok med dansk forklaring i databaseformat. I: Caroline Sandström m.fl. (red): *Perspektiv på leksikografi, grammatik och språkpolitik i Norden*. Helsingfors: Institutet för de inhemska språken och författarna, 81-99.
- Hanks, Patrick (1979): To what extent does a dictionary definition define? I: R. R. K. Hartmann (ed.): *Dictionaries and their Users, papers from the 1978 B.A.A.L. Seminar on Lexicography*. Exeter: University of Exeter, 32-38.
- Haugen, Einar (1972): Analysis of a sound group: SL and Tl in Norwegian. I: Evelyn S. Firchow (red.): *Studies by Einar Haugen, Presented on the Occasion of his 65th Birthday, April 19, 1971*. Berlin & The Hague: de Gruyter, 115-141.
- Kilgariff, Adam (1993): Dictionary Word Sense Distinctions: An enquiry into their nature. I: *Computers and the Humanities* 26 (1-2), 365-387.
- Nichols, Stephen G. (1990): Introduction: Philology in a Manuscript Culture. I: *Speculum* 65, 1-10. doi:10.2307/2864468.
- NO2014 = Prosjektet Norsk Ordbok 2014. <www.norskordbok.uio.no>.
- St.m.35 (2007) = Stortingsmelding 35 (2007-2008) *Mål og mening*.
- TEI (2015) = Text Encoding Initiative, *Guidelines to Text Encoding (P5)*. <www.tei-c.org>.
- UNESCO (2003) = *The Convention for the Safeguarding of the Intangible Cultural Heritage*. UNESCO, MISC/2003/CLT/CH/14, Paris, 2003. <www.unesco.org/culture/ich/en/convention>

Christian-Emil Smith Ore
førsteamanuensis, dr.scient.
Universitetet i Oslo
Institutt for lingvistiske og nordiske studier
Enhet for digital dokumentasjon
Boks 1102, Blindern
NO-0317 Oslo
c.e.s.ore@iln.uio.no

Personbetecknande verbavledningar i Ordbok över Finlands svenska folkmål. En utvärdering av ordbokens sökfunktioner

Caroline Sandström & Carola Åkerlund

The study examines the various search possibilities in the e-dictionary *Ordbok över Finlands svenska folkmål* (Dictionary of Swedish Dialects in Finland) published in 2013. The dictionary project was initiated as early as in the middle of the 20th century, and the concrete editing process started in the 1960s; by 2007 four volumes of the dictionary had been published. To evaluate the search function we examine how a group of derivational suffixes, i.e. *-are, -bla, -dra, -el, -er, -ing, -lla, -ska, -ssa, -us*, are distributed in the e-dictionary, according to their geographical spread and semantic differences. Our results show that the search function developed for the dictionary is a helpful and efficient tool for finding patterns in the distribution of derivational suffixes, but for some purposes the search results need to be further analyzed manually.

1. Inledning

Ordbok över Finlands svenska folkmål (i fortsättningen FO) är en dokumentationsordbok över de svenska dialekter som talas vid finländska väst- och sydkusten. Ordbokens historia sträcker sig till mitten av 1900-talet. Utgivningen inleddes år 1976 och fram till 2007 har fyra tryckta band utkommit¹. Arbetet med att skapa en nätordbok av FO inleddes åren 2010–2011.

Vårt syfte är att studera ett urval personbetecknande verbavledningar i ordbokens material och samtidigt utvärdera sökmöjligheterna i FO som nätordbok. Vi granskar särskilt om det via sökningarna framträder geografiska och semantiska skillnader vid förekomsten av olika avledningssuffix.

¹ För kort bakgrund till ordboksprojektet, se <kaino.kotus.fi/fo> under fliken ”Om ordboken”, Ahlbäck (1976), Slotte (1981), Cantell & Sandström (2012:157f.).

I Cantell & Sandström (2012) diskuteras planering och praktiska problem vid överföringen av den tryckta FO till en nätordbok. På konferensen om lexikografi i Norden 2013, då ett pilotavsnitt (*i-k*) hade publicerats på nätet, diskuterades utmaningar vid lemmatisering och länkning². Det mål som uppställdes var att ordbokens hänvisningar skulle länkas så att man genom att klicka på en länk fördes till det ställe som avsågs (Cantell & Sandström 2012:165). Genom fortsatt tätt samarbete mellan FO:s redaktion och språkteknologerna på Institutet för de inhemska språken kunde målet nås vid uppdateringen i november 2014.

Målsättningen var att FO på nätet skulle bli mångsidigt sökbar, så att det var möjligt att göra enkla sökningar i lemmalistan, avancerade sökningar med avgränsningar och filter och sökningar i ordbokens hela material (Cantell & Sandström 2012:165). Redan vid publiceringen av pilotavsnittet 2013 var FO:s sökfunktioner utbyggda så att sökningarna kunde avgränsas till olika element i artiklarna. Förutom bland uppslagsorden – simplexord och/eller sammansättningar – är det i nätordboken möjligt att söka bland uttalsvarianter och exempelfraser. Genom funktionen ”Sök i hela texten” kan man göra sökningar även i betydelseförklaringarna, trots att dessa inte märkts upp som separata element. Tack vare sökfilter kan man också avgränsa sökningen bl.a. enligt beläggort eller dialektområde, uttalsvariant, ordklass, böjningskategori (deklination och konjugation) och etymologi (t.ex. finska, ryska, amerikaengelska som långgivande språk).

Utifrån Tarp (2012:116f.) kan ambitionerna för FO som nätordbok hänföras till kategori två av de fyra kategorier han iakttar i utvecklingen från pappersordböcker till moderna interaktiva e-ordböcker. Kategorin innefattar en stor grupp traditionella e-ordböcker som enbart i begränsad utsträckning dragit nytta av landvinningar inom lexikografisk teori med moderna lexikografiska verktyg och teknologiska resurser. Hanks (2012:57ff., 82) och Abel (2012:83ff.) understryker att det inom dagens lexikografi finns stora möjligheter tack vare den snabba utvecklingen inom elektronisk lexikografi med tillgång till oändligt stora databaser och effektiva redigeringsprogram. Men samtidigt framhåller de att både

² Sektionsföredrag ”Lemmatisering och länkning i Ordbok över Finlands svenska folkmål på nätet” av Sandström, C. & Outi Lehtinen & Risto Widenius på 12 konferensen om lexikografi i Norden, Oslo 13–16 augusti 2013 (opublicerat).

lexikografer och förlag lever i en hård värld med ekonomiska realiteter och begränsningar.

I detta bidrag utvärderar vi sökmöjligheterna i nätordboken FO för att få en uppfattning om hur man kan hantera dialektmaterial med andra metoder än de traditionella, så som att manuellt excerpera uppteckningar och bandinspelningar. I arbetet med att redigera ordboken har vi haft en känsla av att det finns en viss snedvridning i materialet då det t.ex. gäller förekomsten av pejorativa benämningar för personer och vissa typer av personavledningssuffix. Med hjälp av den avancerade sökfunktionen undersöker vi därför om det finns geografiska och semantiska skillnader i bruket av verbavledningar i FO. Vi har avgränsat oss till avsnittet *a–k* som var publicerat på nätet våren 2015 och undersöker personbetecknande avledningar med suffixen *-are, -bla, -dra, -el, -er, -ing, -lla, -ska, -ssa, -us*. Viktiga frågeställningar är bl.a. om det finns risk för att materialet i nätordboken blir ohanterligt med för många irrelevanta träffar och i vilken utsträckning sökningarna måste kompletteras med manuella kontroller.

2. Verbavledningar i svenskan

I undersökningen fokuserar vi på personbetecknande suffixavledningar. Enligt en indelning i Josefsson (2005:105ff.) hör verbavledningar till gruppen komplexa ord. Gruppen kan indelas i sammansättningar och avledningar, och vidare kan avledningar vara antingen prefix- eller suffixavledningar. Josefsson (2005:117) framhåller att avledningssuffixens funktion är ”att modifiera det som de har förenats med och på så sätt skapa ett nytt ord”.

En grupp bland personbetecknande suffixavledningar som de flesta handböcker lyfter fram är det substantivbildande suffixet *-are*, nomen *agentis* (Josefsson 2005:124). Det är allmängermanskt och produktivt (Thorell 1984:78, Liljestränd 1993:69). Suffixet betecknar personer som yrkes- eller vanemässigt utövar viss verksamhet (Thorell 1981:103). Thorell (1981:105) framhäver också att suffixet *-are* är primärt i förhållande till suffixet *-ska*. Hellquist (1922) påpekar att suffixet *-er*, som i fornsvenska markerade nominativ singular maskulinum, är en kvarleva i ord med

förklenande betydelse – t.ex. FO:s *blasker* eller *burker* ’tölp’ och FO:s *gormer* ’skrikhals’ – och kan ge upphov till nybildningar.

Widmark (1996) diskuterar nedsättande ord i dialekter utifrån Gustaf Ericsson samlingar från Södermanland (slutet av 1800-talet). Hon poängterar att sådana var erhört frekventa bland den tidens allmoge, men ofta enbart lokala till sin utbredning. Orden var inte alltid grovt nedsättande utan kunde användas skämtsamt och sedan spridas. Men det finns en gråzon mellan det klart fördömande och det humoristiska vilket har samband med de normer och värderingar som rådde i samhället; man skulle inte avvika från det normala.

Då de nedsättande benämningarna uppstått har personer som uppfattats som dumma och enfaldiga, otympliga och tafatta, slarviga och orediga, tillgjorda, inställsamma, lömska, framfusiga och bullriga, eller missnöjda och snarstuckna, varit måltavla och benämnts med pejorativer, t.ex. *skrävlare* eller *kloker* ’snusförnuftig person’. Som positiva egenskaper uppfattades däremot egenskaper som att vara vettig, duglig, stillsam och att visa respekt för sina medmänniskor (Widmark 1996:88).

Kvinnor bedömdes huvudsakligen utifrån hur de fungerade i hemmet, medan männen bedömdes enligt sin förmåga att samarbeta och arbeta hårt. Negativa omdömen drabbade kvinnor som uppfattades som otuktiga, slarviga, snuskiga, elaka och trätgiriga, t.ex. FO:s *bubbla* ’ilsken, trätande kvinna’. Männen däremot skulle inte vara tjurskalliga och tvära (t.ex. FO:s *kinkus* ’grinolle, kinkblåsa’), självsvåldiga, småaktiga, fjäskiga eller glupska. I arbetet med vårt finlandssvenska material har vi iakttagit liknande mönster i fördelning av såväl egenskaper som könsfördelning som Widmark gjort, vilket också syns i att det var lätt att hitta exempel i FO:s material.

3. Metod och sökningar

Sökningarna på suffixavledningar gjordes i det avsnitt som våren 2015 fanns tillgängligt av nätordboken, dvs. avsnittet *a–k*. Vi utgick från avledningarna *-are*, *-bla*, *-dra*, *-el*, *-er*, *-ing*, *-lla*, *-ska*, *-ssa*, *-us* som är eller kan vara avledda av ett verb. Vid valet av avledningar utgick vi från frekvens. Mest frekventa var typerna *-are*, *-ska*, *-er*, *-ing* och *-us*, vilka även nämns i referenslitteraturen som produktiva avledningar (Thorell 1981:103

och 1984:78, Liljestrand 1993:69). En kategori med något färre träffar var avledningarna *-bla*, *-dra*, *-el*, *-lla* och *-ssa*. Vissa avledningar utelämnades då verb saknas som grund eller träffarna var alltför få (*-nna*, *-lle*, *-ika*, *-ul*, *-ur* och *-or*).

I en del fall finns det ett annat personbetecknande substantiv med samma betydelse, som då troligen är det ord som avledningen bygger på (Thorell 1981:106). I fråga om suffixet *-er* i FO, t.ex. *blasker* 'slarv, tölp', *fjönter* 'narr, fjant', *köver* 'uppkäftig pojke', kan det vara svårt att bedöma om ursprunget är en synonym personbeteckning (*blask*, *fjönt*, *köv*) eller om avledningen primärt bygger på motsvarande verb (*blaska* 'skämma ut sig', *fjönta* 'bete sig narraktigt', *kövra* 'gnata, tjata').

De konkreta sökningarna gjordes i nätordbokens avancerade sökning (se <kaino.kotus.fi/fo>) genom att i rutan "Sök uppslagsord" skriva in jokertecknet * + suffixet (t.ex. *el). Sökningarna begränsades ytterligare till kategorierna "huvudord" och "substantiv" för att utesluta bl.a. sammanställningar och verb. Vissa sökningar gav många irrelevanta träffar vilket gjorde att manuell kontroll krävdes för att utesluta icke-suffixavledningar.

4. Resultat

Våra sökningar gav som resultat att suffixet *-are* är det vanligaste personbetecknande avledningssuffixet i FO, vilket vittnar om att de svenska dialekterna i Finland ansluter sig till typiska svenska ordbildningsmönster (Thorell 1981:103 och 1984:78, Liljestrand 1993:69, Josefsson 2005:124). Av Tabell 1 framgår att det fanns drygt 350 relevanta träffar med personavledningen *-are*. Andra stora grupper var personavledningarna *-us* och *-el*. Bland de personbetecknande orden med suffixet *-are* finns vissa semantiska skillnader, vilket främst beror på att suffixet är så frekvent och bildar en omfattande och heterogen grupp. En betydande andel av orden med suffixet *-are* avser personer som yrkes- eller vanemässigt utövar viss verksamhet, t.ex. FO:s *basare* 'man som mjukade upp och böjde gärdsgårdsband', *bråkare* 'person som bråkade lin', *junnare* 'flottningsarbetare'³. Men det fanns också en stor grupp personavledningar

³ Personbeteckningen *junnare* är avledd av verbet *junna* 'flotta timmer' som går tillbaka på det dialektala finska verbet *junnata* 'röra sig långsamt'; se FO:s *junna*.

på *-are* som har en tydlig pejorativ betydelse, t.ex. FO:s *flöskare* 'sluskig karl' och *krapplare* 'sjukling; fuskare'.

Suffix	Antal personavledning	Totalt antal träffar
<i>-are</i>	353	661
<i>-us</i>	152	273
<i>-el</i>	95	527
<i>-er</i>	41	280
<i>-ing</i>	29	476
<i>-ska</i>	15	118
<i>-lla</i>	9	67
<i>-bla</i>	7	19
<i>-dra</i>	6	21
<i>-ssa</i>	3	47
Summa:	710	2 489

Tabell 1: Resultat för sökningarna av de tio suffixavledningarna i FO.

Som framgår av Tabell 1 var de relevanta sökresultaten för vissa suffix ganska få. Exempelvis gav sökningen av ord som slutar på *-ing* 29 relevanta träffar medan sökning på suffixet *-el* gav totalt 527 träffar av vilka 95 visade sig vara personavledning av ett verb.

Våra resultat visar att även betydelsemässigt neutrala personbeteckningar för både män och kvinnor finns, som t.ex. *horpare* 'person som går kippskodd', *gryckel* 'stelbent och stapplande människa', *irvus* 'person som grimaserar'.

Personsuffixet *-ska* betecknar kvinnor som utövar viss verksamhet, t.ex. *bykerska* 'tvätterska', *gniderska* 'äldre kvinna som masserar för muskelsträckning', medan suffixen *-bla*, *-dra*, *-lla*, *-ssa* vanligen används pejorativt för kvinnor: *jobbla* 'sladdrande kvinna', *fluddra* 'lättsinnig kvinna', *dulla* 'enfaldig, oföretagsam kvinna', *dussa* 'kvinna som stökar omkring utan att få något uträttat'.

Sökningarna visade att det finns geografiska skillnader i fördelningen av personsuffixen. Dessa mönster gäller även suffix med pejorativ innebörd. Suffixet *-el* har viss dominans i Österbotten t.ex. *dackel* 'dagdrivare; lätting', *kryckel* 'ynkrygg', medan *-er* förekommer mer i södra Finland (Åland, Åboland, Nyland) t.ex. *dröler* 'sölare', *gluffer* 'storätare; tölp'. Även suffixet *-us* dominerar i södra Finland (Åboland och västra Nyland), t.ex. *daskus* 'slaver', *kräpplus* 'grälsjuk man'. De geografiska skillnader som framträder i våra sökresultat var inte en överraskning eftersom vi i arbetet med materialet till ordboken iakttagit att insamlare på vissa orter verkar ha varit särskilt fokuserade på pejorativa benämningar. På dessa orter är också vissa pejorativa suffix mer frekventa, t.ex. suffixet *-us* i Nagu och Pargas.

Som ovan konstaterats har vi i FO lagt märke till liknande mönster som finns hos Widmark (1996). Varken kvinnor eller män skulle vara alltför pratsamma eller skvallriga, t.ex. FO:s *bibla* 'sladdertaska' och FO:s *bralla* 'skrävare' är exempel på pejorativa personavledning för sådana personer. För båda könen uppfattades också lättja och sysslolöshet som negativt, t.ex. FO:s *dalla* 'kvinna som går lat och sysslolös' och FO:s *döding* 'lätting som bara vill masa sig'.

5. Utvärdering

Fördelen med sökningar i en nätordbok jämfört med traditionella analysmetoder är att man kan hantera stora material och att man hittar belägg som man kanske hade missat med manuella metoder (jfr även Abel 2012, Hanks 2012). Dessutom är analysen mindre beroende av forskarens förhandsuppfattningar, dvs. man upptäcker mönster och tendenser, så som geografiska mönster i fördelningen av suffixtyper.

Sökningar i större textmassor i den redigerade ordboken gör att man inte har samma omedelbara kontakt med källmaterialet som då man arbetar mer traditionellt med upptecknat arkivmaterial eller bandinspelningar. (Jämför även diskussionen i Hovmark 2012:301f., 305f.). I den aktuella nätversionen kan sökningarna inte avgränsas tillräckligt utan ger mycket onödigt brus, vilket innebär att man måste göra kompletterande kontroller och analyser.

Ett alternativ hade varit att använda funktionen "Sök i hela texten", dvs. söka via betydelseförklaringarna i brödtextern, men eftersom redaktörerna vid

redigeringen av FO haft fria händer kan förklaringarna i vissa fall bygga på individuella lösningar och vara utformade efter varierande mönster, vilket gör att det kan vara svårt att komma på sökord som ger adekvata resultat.

Litteratur

Ordböcker

FO = *Ordbok över Finlands svenska folkmål*. (2013–). Institutet för de inhemska språkens webbpublikationer 33. Helsingfors: Institutet för de inhemska språken. <kaino.kotus.fi/fo> (april 2015).

Annan litteratur

Abel, Andrea (2012): Dictionary writing systems and beyond. I: *Electronic lexicography*. Oxford: University Press, 83–106.

Ahlbäck, Olav (1976): Inledning. I: *Ordbok över Finlands svenska folkmål*. Helsingfors: Forskningscentralen för de inhemska språken, VII–XXVIII.

Cantell, Ilse & Caroline Sandström (2012): Hur blir en traditionell, tryckt ordbok en webbordbok? I: *Nordiska studier i lexikografi 11. Rapport från Konferensen om lexikografi i Norden. Lund 24–27 maj 2011*. Lund, 157–168.

Hanks, Patrick (2012): Corpus evidence and electronic lexicography. I: *Electronic lexicography*. Oxford: University Press, 57–82.

Hellquist, Elof (1922): *Svensk ordbildningslära från historisk synvinkel*. Lund: Gleerups.

Hovmark, Henrik (2012): Data og repræsentativitet i ordbogsarbeidet. I: *Nordiska studier i lexikografi 11. Rapport från Konferensen om lexikografi i Norden. Lund 24–27 maj 2011*. Lund, 296–308.

Josefsson, Gunlög (2005): *Ord*. Lund: Studentlitteratur.

Liljestrand, Birger (1993): *Så bildas orden*. Lund: Studentlitteratur.

Slotte, Peter (1981): Ordbok över Finlands svenska folkmål – problem och metoder. I: *De finlandssvenska dialekterna i forskning och funktion*. Åbo: Åbo Akademi, 45–64.

- Tarp, Sven (2012): Theoretical challenges in the transition from lexicographical p-works to e-tools. I: *Electronic lexicography*. Oxford: University Press, 107–118.
- Thorell, Olof (1981): *Svensk ordbildningslära*. Stockholm: Esselte studium.
- Thorell, Olof (1984): *Att bilda ord*. Malmö: Scriptor Förlag.
- Widmark, Gun (1996): Drög och lortlolla. Om nedsättande ord i dialekten. I: *Svenska landsmål och svenskt folkliv 1996*. Uppsala, 87–92.

Caroline Sandström
huvudredaktör, fil. dr.
Ordbok över Finlands svenska folkmål
Institutet för de inhemska språken
Hagnäskajen 6
FI-00530 Helsingfors
caroline.sandstrom@sprakinstitutet.fi

Carola Åkerlund
redaktör, fil. mag.
Ordbok över Finlands svenska folkmål
Institutet för de inhemska språken
Hagnäskajen 6
FI-00530 Helsingfors
carola.akerlund@sprakinstitutet.fi

Den mobile ordbogsbruger

Henrik Køhler Simonsen

People are already mobile – but is lexicography as a discipline up for the mobile challenge? This article is based on empirical data from a survey of five teenagers, who were asked to look up five terms with a bilingual dictionary app. The empirical data showed how real teenage users interact with mobile phones and dictionary apps and seem to indicate that lexicography and mobile devices are not yet fully compatible in all situations.

1. Indledning og problem

Leksikografien har været på en fantastisk rejse de sidste 20 år. Fra papir til internet og fra internet til smartphone. Denne gang er der dog ikke kun tale om et teknologiskifte, men et decideret paradigmeskifte, som på mange måder har været disruptiv, idet ikke kun mediet, men også brugeren, behovet, brugssituationen, tilgangen, designet og indholdet har ændret sig. IT-generationen stiller nemlig store krav til både tilgang og indhold, som det fremgår af Prenskey (2001).

Brugerne er i den grad allerede mobile, og brugerne bruger i stadig stigende grad mobiltelefonen til flere og flere komplekse opgaver, som det fremgår af f.eks. Google (2013). Men brugerne har ikke samme betalingsvilje som tidligere og en udbredt ”good enough-holdning” præger ofte IT-generationen. Mobiltelefonen har desuden også stadig typisk kun en 7-tommer skærm, og man kan som bruger ikke tilgå ret mange data på så lille en skærm – og da slet ikke så mange data, som der ofte findes i traditionelle ordbogsartikler.

Ifølge Nielsen (2011) er netop skærmens størrelse en kognitiv udfordring, fordi det simpelthen er sværere at uddrage og bruge den information som man har brug for. Nielsen (2011) argumenterer endvidere, at brugerne de facto er utålmodige, og at de kognitivt ikke kan overskue så mange data på så lille en skærm. Der findes allerede mange løsninger, der gør ordbogsapps mere overskuelige, f.eks. Expand-Collapse-løsninger eller side- og top-menuer, men spørgsmålet er om disse løsninger er nok? Måske skal nogle af

leksikografiens grundprincipper gentænkes eller måske skal leksikografiens høje kvantitetsmæssige kvalitetsparametre gentænkes? Den kætterske tanke om, at der på mange områder skal præsenteres færre og mere enkle data i mobileksikografi ligger lige for.

Så hvorfor bliver leksikografien ved med at præsentere alt for mange data på alt for lidt plads? Dette spørgsmål diskuteres bl.a. af Tarp (2015:17), som rammende argumenterer, at mange e-ordbøger desværre stadig fremstår som elektroniske udgaver af papirordbøger med traditionelle ordbogsartikler. Den tendens ses desværre også i ordbogsapps. Der er masser af eksempler på ordbogsapps, som slet ikke udnytter alle de muligheder det mobile medie har, hvilket for så vidt også er en del af udfordringen.

Formålet med denne artikel er således først at diskutere en ny metode til indhentning af brugerdata om digitale ordbogsbrugeres brugeradfærd samt dernæst at diskutere, hvordan den unge digitale ordbogsbruger anvender en udvalgt digital ordbog på en smartphone. Med udgangspunkt i en mindre, prototypisk empirisk undersøgelse af fem teenagere, viser artiklen, hvordan unge ordbogsbrugere bruger deres smartphone til at tilgå ordbogsdata, og hvordan de synes at tilegne sig sproglig viden. De unge digitale ordbogsbrugere synes i særlig grad at anvende deres smartphone hele tiden, også når de bevæger sig. Derfor gennemføres der to forsøg – et hvor de digitale ordbogsbrugere sidder stille og et hvor de bevæger sig. På basis af disse indsigter vil artiklen diskutere de unge digitale ordbogsbrugeres nye behov og brugssituationer, og hvad dette betyder for både design og indhold i moderne ordbogsapps.

2. Empiri og metode

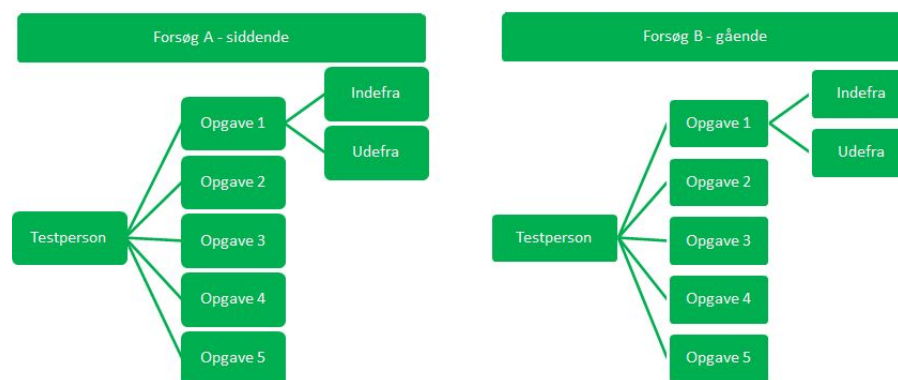
Den undersøgelse, som danner basis for diskussionen af mobileksikografi i denne artikel, blev gennemført i efteråret 2014 og omfattede fem 13-årige teenagere. De fem teenagere blev bedt om at slå fem almensproglige termer op fra en engelsk tekst, som i 2013 blev anvendt i en nationaltest i engelsk i 7. klasse. Forsøget blev udført med ordbogsappen Gyldendal Engelsk-Dansk, som blev tilgået fra en iPhone 4S. For at kunne optage den faktiske brugeradfærd var mobiltelefonen forbundet trådløst med Reflector, som anvendes til at optage en film af brugernes interaktion med telefonen. De fem

forsøgspersoner blev også optaget udefra vha. et almindeligt digitalkamera, således det blev muligt at analysere, hvordan de begår sig i det fysiske rum samtidig med, at de bruger mobiltelefonen. Endelig blev der optaget tænkehøjt-data under hele forløbet.

Der blev gennemført to forsøg, hvor forsøgspersonerne først blev bedt om at slå de fem termer op, mens de sad ved et bord, og derefter blev de bedt om at gøre det samme, mens de langsomt gik omkring et bord. Tilgang til data skerr i stadig stigende grad under en eller anden form for bevægelse, hvilket i særlig grad synes at være tilfældet for unge teenagere. Derfor gennemførtes to forsøg med henblik på at kunne identificere brugernes adfærd i mobile brugssituationer. Forsøgspersonerne blev bedt om at slå *wildlife programmes*, *cheetahs*, *fancy it*, *auntie* og *disappointed* op med henblik på at oversætte dem til dansk.

Forsøgene testede således, hvordan de fem forsøgspersoner agerer i en kommunikativ situation – siddende og gående – i den intra-leksikografiske konsultationsfase og i den extra-leksikografiske post-konsultative fase, som det fremgår af Tarp (2015:20).

En oversigt over de empiriske data fremgår af Figur 1 herunder.



Figur 1: Oversigt over empiriske data.

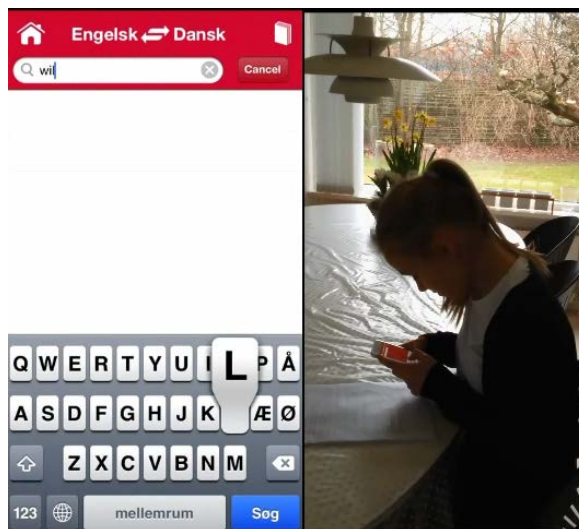
De fem 13-årige går på Birkerød Skole, som er en almindelig dansk folkeskole. Skolen har udleveret iPads til alle elever, hvilket betyder, at de fem forsøgspersoner er meget teknologikyndige og vant til at arbejde med den pågældende app via skolens UniLogin. De fem forsøgspersoner har i forvejen selv en iPhone og kender dermed dens begrænsninger og muligheder.

3. Analyse og diskussion

Den første forudsætning for at diskutere mobileksikografi er at diskutere brugerundersøgelser og beskrive den anvendte metode. På den ene side hævder Tarp (2008:44), at brugerundersøgelser af specifikke leksikografiske situationer er som at "fylde Danaidernes kar", altså noget nytteløst. På den anden side hævder f.eks. Nielsen (2000) og Lew (2015) m.fl., at brugerundersøgelser er helt nødvendige, og at empiriske data er forudsætningen for at udvikle bedre værktøjer på et videnskabeligt grundlag.

Lew (2015:5-9) f.eks. diskuterer de fundamentale forskelle i hhv. det positivistiske paradigme, som fokuserer på kausalitet og generaliserbarhed af resultater, dvs. den typiske fremgangsmåde inden for naturvidenskab, og det naturalistiske paradigme, som tager udgangspunkt i kvalitative data og fokuserer på fortolkninger af disse.

Denne undersøgelse tilhører ikke overraskende det naturalistiske paradigme, og det er som det fremgår en mindre prototypisk undersøgelse. Både de kvantitative og kvalitative data, som undersøgelsen har tilvejebragt, er blevet indsamlet og fortolket i den situation, de hører til. Det er bl.a. derfor, at de to videooptagelser både indefra og udefra er blevet synkroniseret, således man i den sammenklippede film kan se, hvordan brugeren interagerer med mobiltelefonen og ordbogsappen samtidig med at man kan se, hvordan brugeren interagerer med virkeligheden og det fysiske rum. Det er i særlig grad interessant at se, hvordan ordbogsbrugers kognitive investering i navigation i det fysiske rum, f.eks. ved hele tiden at se op fra telefonen og ved at undgå at gå ind i objekter, direkte påvirker den måde de interagerer med ordbogsappen på, samt på den måde de ekstraherer og behandler data på. Et eksempel på, hvordan de sammenklippede film ser ud, vises i Figur 2.



Figur 2: Sammenklippet film – forsøgsperson 3 – indefra og udefra.

På basis af denne og flere andre undersøgelser, se også Simonsen (2015), argumenteres det, at den udviklede undersøgelsesmetode er velegnet til leksikografiske brugerundersøgelser, og at metoden kan afsløre, hvordan den digitale ordbogsbruger rent faktisk anvender en ordbogsapp. Selve undersøgelsen her omfatter ganske vist kun fem unge forsøgspersoner, men ifølge Nielsen (2000) er det nok til at kunne få indsigt i, hvordan en lille gruppe af forsøgspersoner anvender et digitalt testobjekt. Nielsen (2000) anfører f.eks. at "Elaborate usability tests are a waste of resources. The best results come from testing no more than 5 users and running as many small tests as you can afford" på basis af $N(1-(1-L)^n)$, hvor N er det totale antal iagttagelser og L er andelen af iagttagelser ved test med en enkelt person. Ifølge Nielsen (2000) er den typiske værdi af L 31 % hvilket betyder, at man allerede ved fem testpersoner får et forholdsvist stort testudbytte, og at udbyttet for hver testperson derefter er marginalt. På trods af Nielsens argumenter her, anføres det dog, at det i høj grad ville være relevant at supplere denne mindre prototypiske undersøgelse med langt flere testpersoner, herunder flere forskellige typer brugere. En gruppe på fem brugere er således ikke repræsentativ for alle typer af brugere, og som det senere bliver diskuteret, synes der også at være flere forskelle mellem teenagere og f.eks. læger, jf. Simonsen (2015).

Den anden forudsætning for at diskutere mobilleksikografi er at tage udgangspunkt i resultaterne fra empiriske undersøgelser og se på, hvordan rigtige brugere anvender mobiltelefonen i konkrete brugersituationer til at løse konkrete opgaver.

Undersøgelsen testede som bekendt, hvordan de fem 13-årige slog fem opslagsord op siddende og gående. Undersøgelsen bekræfter ikke overraskende, at brugerne helt naturligt synes at taste hurtigere og mere sikkert, når de sidder ned, end når de bevæger sig. Undersøgelsen påviste også, at forsøgspersonerne synes at kunne tilgå data hurtigere, når de sidder ned, end når de bevæger sig, dvs. de er bedre til at finde de ønskede data i den stationære brugssituation. Undersøgelsen viste også, at når de så bevæger sig og samtidig interagerer med mobiltelefonen, så går begge dele meget langsomt – dvs. de bevæger sig langsomt, fordi de skal orientere sig i både skærm og i det fysiske rum på samme tid. Brugere synes at holde mange pauser ved indtastningen, mens de går rundt og samtidig er der en tendens til, at de også fysisk holder pause og står stille. Med andre ord så er den bevægelige brugssituation ikke på nogen måde befordrende for nogen af de tre konsultationsfaser, som diskuteret af Tarp (2015:20).

Undersøgelsen testede også, hvordan de fem 13-årige rent faktisk interagerede med mobiltelefonen og ordbogsappen. For det første blev det hurtigt klart, at brugerne oftest bruger tohåndsgreb, og at de taster med begge tommelfingre. Kun en ud af fem forsøgspersoner brugte mobiltelefonen i horisontal stilling, hvilket for så vidt er en smule overraskende, idet netop læsefeltet bliver væsentligt bredere ved horisontal placering. De mange optagelser påviste også klart, at forsøgspersonerne meget, meget ofte taster forkert, og at de bruger Backspace-knappen for at slette og så taste igen. Mobiltelefonens Autofill-funktion synes at være en stor hæmsko for brugerne, fordi den hele tiden var et forstyrrende element frem for en aktiv hjælp. Tre ud af de fem testpersoner brugte endda forstørrelsesglasset, når de rettede stave- eller slåfejl.

Optagelserne både indefra og udefra og de indsamlede tænke-højt-data synes også at vise, at de fem 13-årige relativt hurtigt synes at give op, hvis de ikke fandt lige præcis det de søgte efter, ligesom de meget hurtigt gætter på oversættelsen, også selvom ordbogsappen på ingen måde har givet dem det svar, de var ude efter. Dette adskiller sig meget fra den gruppe testpersoner, som Simonsen (2015) diskuterer. Denne testgruppe bestod af 10 læger,

som blev bedt om at slå fem navne på lægemidler op. Denne gruppes brugeradfærd adskiller sig fra de fem 13-åriges brugeradfærd ved, at lægerne er langt mere grundige og bliver ved, indtil de har fundet et svar, de er tilfredse med. Dette synes på ingen måde at være tilfældet for de fem 13-årige, som meget hurtigt giver op og gætter. Der synes således at være en klar forskel på, hvordan ”digital natives” og ”digital immigrants” agerer, som diskuteret af Prensky (2001). Denne markante forskel mellem de unge og de ældre forsøgspersoner kan skyldes flere ting. For det første anføres det, at mange unge ikke i tilstrækkelig grad har lært, hvordan man bruger mobilordbøger eller ordbøger i det hele taget. For det andet anføres det, at selvom den yngre generation synes at være ekstremt digitale og stort set online hele tiden, så synes de ikke i tilstrækkelig grad at have lært at sortere i og vælge de mange data, som ordbogsappen præsenterer for dem. Andre undersøgelser har påvist noget lignende, f.eks. især Curcio (2014) men også Marelló (2014) og Müller-Spitzer (2008), som alle blandt andet diskuterer unge ordbogsbrugeres ringe reference skills.

At den yngre generation ikke i tilstrækkelig grad synes at have lært, hvordan man anvender opslagsværker, spiller således også en særdeles vigtig rolle ved design af mobile opslagsværker. Den yngre generation af ordbogsbrugere har formentlig ikke i samme grad som den lidt ældre generation lært at afkode og forstå de mange etablerede konventioner og standarder i leksikografien. Dette betyder for det første, at en mobilordbog, som bruger designstandarder og konventioner fra papirmediet vil fremstå antikvarisk og vanskelig at bruge for den unge bruger og for det andet, hvilket i og for sig er det vigtigste, at en mobilordbog kan og bør designes ud fra sine helt egne præmisser og ikke mindst på basis af indgående viden om hvordan den unge ordbogsbruger anvender ordbogsappen – i siddende, stationære brugssituationer såvel som i gående, mobile brugssituationer. Vi er således lige nu i en helt unik situation, fordi vi kan udvikle nye leksikografiske produkter på basis af indgående viden om de unges brugeradfærd. Men for at give den unge ordbogsbruger maksimal udbytte af mobilordbogen og dermed mulighed for at realisere det fulde potentiale i mobilordbøger skal der gennemføres langt flere studier af de unges brugeradfærd og af hvordan de tilgår f.eks. flerordstermer. Der er behov for at finde ud af, hvorvidt de unge ordbogsbrugere foretrækker aktiv eller passiv brugeradfærd, og der er behov for

at vide, om mobile ordbogsbrugere overhovedet orker at svare på hjælpspørgsmål som ”Mente du xxx?”.

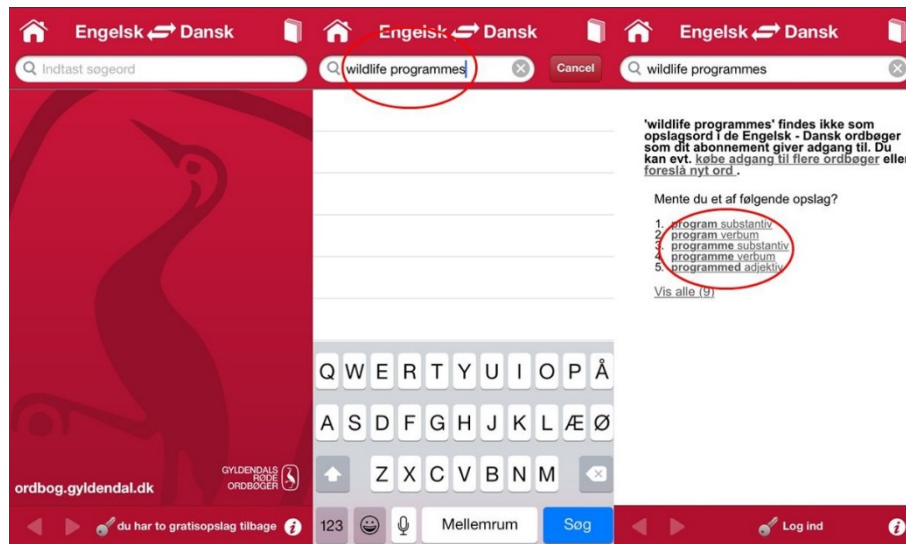
Undersøgelsen påviste, at forsøgspersonerne rent faktisk fandt anvendelige data og foreslog relevante oversættelser i de tilfælde, hvor opgaven var at slå et enkelt ord op. Fem ud af fem forsøgspersoner fik meget hjælp fra ordbogsappen da de slog *cheetahs*, *auntie* og *disappointed* op, og de kom alle fem til tilfredsstillende oversættelser til dansk.

Undersøgelsen viste også klart, at forsøgspersonerne ikke fik ret meget hjælp fra ordbogsappen, da de slog flerordstermerne *wildlife programmes* og *fancy it* op. Kun en ud af de fem forsøgspersoner fandt frem til et tilfredsstillende forslag til oversættelse af de to flerordstermer. Hvorvidt det skyldes de manglende reference skills hos de fem testpersoner eller det skyldes søgemaskinens manglende evne til at forudsige/tolke hvad brugeren reelt søger efter i den pågældende app kan naturligvis diskuteres. Undersøgelsen synes at pege på begge dele. De fem 13-årige indtastede typisk begge ord i søgefeltet og søgte, og selvom ordbogsappen i bedste leksikografiske stil viser et for papirordbøger overskueligt indeks, så var der ingen af de fire forsøgspersoner, som lagde mærke til det, eller brugte den hjælp de blev tilbudt. Forsøgsperson 4 udtalte *Det kan jeg godt nok ikke finde ud af* og forsøgsperson 2 sagde *Den kan ikke finde det* eller som forsøgsperson 3 udtaler *Den kommer heller ikke op*.

Men denne situation skal den unge mobilbruger slet ikke havne i. En mere intelligent søgemaskine, som man f.eks. ser det i appen Medicin.dk, er at foretrække til det mobile medie. En søgemaskine, som tolker brugerens indtastning intelligent og som automatisk foreslår anden del af flerordstermen er klart den bedste løsning. Undersøgelsen her viser jo med al tydelighed, at mobilbrugeren aldrig kommer ind til det ellers udmærkede indeks, og så bliver det fulde leksikografiske potentiale aldrig forløst. Til stor skade for alle parter.

De unge digitale ordbogsbrugeres tålmodighedstærskel er ret lav, idet de gav op, hvis ikke ordbogsappen viste et svar de kunne bruge. En anden og måske meget mere nærliggende forklaring kan være, at forsøgspersonerne slet ikke har lært at bruge hverken en ordbog eller en mobilordbog, hvilket for så vidt er endnu mere alvorligt for folkeskolen og de leksikografer, som udvikler løsninger til skolerne.

Et typisk eksempel på en situation, hvor fire ud af fem forsøgspersoner gav op eller gættede vises i Figur 3 herunder.



Figur 3: Opslag af flerordsterm.

Undersøgelsen viser, at forsøgspersonerne slet ikke når ned til den hjælp, som ordbogsappen tilbyder, og ingen af dem lægger mærke til den nummererede indeksliste med ordklasseangivelser. Den aktive stillingtagen som brugeren afkræves er ikke hensigtsmæssig. ”Mente du et af følgende opslag?”-formuleringen synes på basis af denne undersøgelse ikke at virke. Det er muligt, at den virker for f.eks. meget fagligt kompetente voksne, f.eks. læger, som diskuteret af Simonsen (2015), men for unge mobilbrugere i Folkeskolen synes den at være uhensigtsmæssig.

Det er således et spørgsmål om aktiv eller passiv søgetilgang. Skal brugeren afkræves et svar og gives mulighed for at tilrette/tilpasse sin søgning aktivt ved at klikke på skærmen eller skal det foregå automatisk i baggrunden på basis af en intelligent søgemaskine. Undersøgelsen her synes klart at vise, at det sidste er tilfældet.

4. Konklusion

Denne artikel har præsenteret en ny innovativ metode til indhentning af brugerdata. Metoden indebærer, at man får dyb indsigt i brugernes interaktion med mobiltelefonen og ordbogsappen og dermed også en viden om, hvad brugerne rent faktisk kan og gør med de leksikografiske data.

Undersøgelsen af de fem 13-årige synes at pege på i hvert fald tre vigtige konklusioner, som for så vidt sikkert også gælder for andre typer af brugere. For det første bør den unge mobilbruger tilbydes langt bedre søgemaskiner i ordbogsapps end indeks med ækvivalentlister. En intelligent søgemaskine, som ikke afkræver mobilbrugeren aktiv stillingtagen synes at være en tilrådelig løsning. For det andet skal der arbejdes med at undgå ”relative overload”, som er når der er flere data, end man visuelt kan overskue, jf. Tarp (2015:17). Mobilbrugeren kan sagtens få adgang til alle tilgængelige leksikografiske data, men der anbefales Expand-Collapse-løsninger, som gør, at brugeren kan overskue kompleksiteten på den lille skærm. For det tredje bør mobilbrugeren kunne tilbydes en ordbogsapp, som automatisk skifter udseende mellem stationær og bevægelig visning, ligesom det allerede foregår i de fleste sundhedsapps. Ved bevægelig visning opstår let *relative overload* og ordbogsappen skal derfor under bevægelse tilbyde mobilbrugeren en langt mere simpel brugergrænseflade med ganske få knapper og stor skriftstørrelse. Flere undersøgelser med flere forskellige testpersoner og forskellige former for A/B tests er dog nødvendige for at få endnu mere viden om den digitale ordbogsbruger.

Litteratur

- Curcio, Martina Nied (2014): Die Benutzung von Smartphones im Fremdsprachenerwerb und -unterricht. I: *Proceedings of the XVI EURALEX International Congress: The User in Focus*. 15-19 July 2014, Bolzano/Bozen. <www.eurac.edu/en/research/autonomies/commul/Publications/Pages/default.aspx> (1. sep. 2015).
- Google (2013): Our Mobile Planet: Denmark – Understanding the Mobile Consumer. <services.google.com/fh/files/misc/omp-2013-dk-en.pdf> (1. sep. 2015).

- Lew, Robert (2015): Opportunities and limitations of user studies. I: Carole Tiberius & Carolin Müller-Spitzer (Eds.): *Research into dictionary use 5*. Arbeitsbericht des wissenschaftlichen Netzwerks „Internet-lexikografie“. Mannheim: Institut für Deutsche Sprache. OPAL <multimedia.ids-mannheim.de/mediawiki/web/images/7/7f/Preprint-V1.pdf> (1. sep. 2015).
- Marello, Carla (2014): Using Mobile Bilingual Dictionaries in an EFL Class. I: *Proceedings of the XVI EURALEX International Congress: The User in Focus*. 15-19 July 2014, Bolzano/Bozen. <www.eurac.edu/en/research/autonomies/commul/Publications/Pages/default.aspx> (1. sep. 2015).
- Müller-Spitzer, Caroline (2008): Research on Dictionary Use and the Development of User-Adapted Views. I: A. Storrer, A. Geyken, A. Siebert, & K.-M. Würzner (eds.): *Text Resources and Lexical Knowledge*. Selected Papers from the 9th Conference on Natural Language Processing KONVENS 2008. Berlin: de Gruyter, 223-238.
- Nielsen, J. (2000): Why You Only Need to Test With 5 Users. I: NN/g Nielsen Norman Group: <www.nngroup.com/articles/why-you-only-need-to-test-with-5-users> (1.sep. 2015).
- Nielsen, J. (2011): When in doubt, leave it out. I: NN/g Nielsen Norman Group: <www.nngroup.com/articles/condense-mobile-content/> (1. sep. 2015).
- Prensky, M. (2001): Digital natives, digital immigrants part 1. I: *On the Horizon* 9(5), 1-6: <www.emeraldinsight.com/ journals.htm?issn=1074-8121 1> (1. sep. 2015).
- Tarp, S. (2008): *Lexicography in the borderland between knowledge and non-knowledge: general lexicographical theory with particular focus on learner's lexicography*. Berlin/New York: de Gruyter.
- Tarp, Sven (2015): Detecting user needs for new online dictionary projects: Business as usual, user research or...? I: Carole Tiberius & Carolin Müller-Spitzer (Eds.): *Research into dictionary use 5*. Arbeitsbericht des wissenschaftlichen Netzwerks „Internetlexikografie“. Mannheim: Institut für Deutsche Sprache. OPAL <multimedia.ids-mannheim.de/mediawiki/web/images/7/7f/Preprint-V1.pdf> (1. sep. 2015)

Simonsen, Henrik Køhler (2015): Mobile Lexicography: Let's Do it Right This Time! I: *Electronic lexicography in the 21st century: linking lexical data in the digital age*. Proceedings of the eLex 2015 conference, 11-13 August 2015, Herstmonceux Castle, United Kingdom. Ljubljana/Brighton: Trojina, Institute for Applied Slovene Studies/Lexical Computing Ltd., 84-104.

Henrik Køhler Simonsen
ekstern lektor, ph.d.
Copenhagen Business School
Dalgas Have 15
DK-2000 Frederiksberg
hks.ibr@cbs.dk

Nyt liv til støvede ordbøger

Thomas Troelsgård & Nicolai Hartvig Sørensen

In the current age of digitisation, an increasing number of older dictionaries are accessible to dictionary users. Often these sources escape the attention of the modern user, although they could be valuable tools in many situations. There can be several reasons for this: complicated access, poor readability (e.g. a complex entry structure that might work as a deterrent), or just pure coincidence. The Society for Danish Language and Literature has digitised and published a series of historic dictionaries. In this paper, we briefly describe these resources, and we give an account of our attempt to promote them via social media and by connecting the dictionaries to the more popular website of the modern Danish dictionary, Den Danske Ordbog. We argue for a network solution where a lookup is performed within a specific dictionary, but with direct links to other relevant dictionaries added. Finally, we outline our plans for future enhancement of the links through manually curated lemma linking across dictionaries.

1. Indledning

Der findes mange gode ordbøger. Men der findes også mange gode ordbøger, som ikke opnår den udbredelse blandt sprogbrugerne, som de fortjener. Det kan være på grund af tilfældigheder, manglende tilgængelighed, eller fordi ordbøgerne er fra en tid hvor ordbøgers udtryk væsentligt adskiller sig fra moderne ordbøger og dermed kan virke afskrækkende.

I nærværende artikel vil vi fokusere på fire af de historiske ordbøger Det Danske Sprog- og Litteraturselskab (DSL) har digitaliseret og offentliggjort, og vi vil skitsere hvordan vi planlægger at sammenkæde disse ordbøger med hinanden – og med DSL's ordbog over moderne dansk, Den Danske Ordbog (DDO).

2. DSL's historiske ordbøger

DSL har digitalt publiceret en række ældre og nyere ordbøger og ordlister. Fire af disse er relevante for denne artikel og beskrives kort i det følgende.

Tabel 1 viser ordbøgernes omtrentlige periode, tilblivelsesår, omfang og publicering, sammenholdt med DDO.

Ordbog	Periode	Opslagsord	Trykt	På nettet
Moths Ordbog	1680-1715	100.000	Aldrig	2013
Holbergordbog	1700-1750	35.000	1981-1988	2011
ODS + ODS-S	1700-1950	225.000	1918-2005	2005
Meyers Fremmedordbog	1800-1900	80.000	1837...	2014
DDO	1950-	90.000	2003-2005	2009

Tabel 1: Nøgletal for DSL's historiske ordbøger og DDO.

2.1. Moths Ordbog

Ordbogen (jf. Eegholm-Pedersen 2013) beskriver dansk omkring år 1700, og udmærker sig ved ikke at fokusere på det “fine”, litterære sprog, men bredt, også med inddragelse af talesprog og dialektale ord og udtryk. Betydningerne beskrives som hovedregel dels ved danske betydningsbeskrivelser (eller synonymmer), dels ved latinske ækvivalenter.

Ordbogen blev redigeret af lægen og embedsmanden Matthias Moth i perioden ca. 1686-1715, men blev aldrig trykt. Den er opdelt i en leksikon- og en ordbogsdel; i den digitale udgave på mothsordbog.dk er disse dele samlet i samme grænseflade. Moths Ordbog rummer ca. 100.000 artikler (omkring 65.000 fulde artikler, dertil listeord og henvisninger).

Det at der ikke har været redigeret i manuskriptet i de 300 år der er gået fra færdiggørelsen til udgivelsen på nettet, gør Moths Ordbog til en tidsmaskine der bringer os tilbage til årene omkring år 1700. Som læser får man således indblik i forskellige dele af hverdagslivet, især i de områder der optog Moth, som når man, under *kirse*, får gode råd om dyrkning af kirsebær, eller til helbredelse af f.eks. *hidsigbyld*, *nyresvær* eller *dødt kød*.

2.2. Holbergordbog

Ordbogen beskriver Holbergs danske ordforråd, og medtager i et vist omfang desuden sted- og personnavne fra hans fiktionstekster.

Ordbogen blev redigeret fra 1946 som et fælles dansk-norsk projekt, fra 1955 dog udelukkende i Danmark, som et DSL-projekt. Den trykte bog udkom i årene 1981-88 og rummer ca. 35.000 artikler.

Udover at være et nyttigt værktøj ved læsning af Holbergs tekster, giver ordbogen naturligvis indblik i dansk sprogbrug generelt i første halvdel af 1700-tallet.

2.3. Ordbog over det danske Sprog (ODS) med supplement (ODS-S)

Ordbogen beskriver dansk i perioden 1700-1950.

ODS er baseret på Verner Dahlerups forarbejde fra 1882 og frem og blev redigeret på DSL i årene 1915-1956 (udgivet 1918-56), ODS-S i årene 1956-2005 (udgivet 1992-2005). Der er ca. 225.000 opslagsord, heraf 35.000 i ODS-S.

ODS rummer mange ord der stadig bruges i dansk, men som på grund af lav frekvens ikke er medtaget i DDO. Udover at kunne bruges som hjælpemiddel ved læsning af tekster fra perioden, udgør ODS/ODS-S således et glimrende supplement til moderne ordbøger.

2.4. Meyers Fremmedordbog

Ordbogen udkom første gang i 1837, og er siden kommet i flere oplag og udgaver. Seneste udgave er 8. udgave fra 1924, der flere gange er kommet i fotografisk optryk, senest i 1998.

Ordbogen rummer mere end 80.000 opslagsord, bl.a. mange fremmedord der ikke findes i moderne ordbøger; Meyers Fremmedordbog er således nyttig ved læsning af fagtekster fra 1800-tallet og frem til i dag.

3. Hvordan udbreder vi kendskabet til ordbøgerne?

Vi har således på ordnet.dk-platformen fire historiske ordbogsressurser, der også i dag har mange anvendelsesmuligheder, både for specialisten og for den almindelige ordbogsbruger, men som ikke bliver brugt særlig meget. Herudover rummer ordnet.dk jo også DDO, som har mange daglige brugere, og som efterhånden må siges at være kendt langt ud over fagkredse. Hvordan

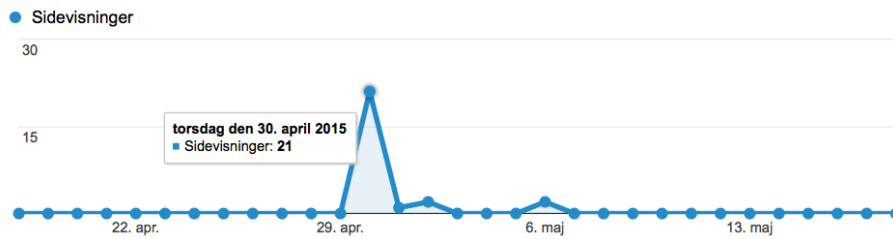
kan vi på denne baggrund hjælpe brugerne til at opdage at de ”hemmelige”, historiske ordbøger findes og faktisk er relevante i mange sammenhænge?

En måde er at promovere ordbøgerne i de sociale medier. Til dette formål har vi på DDO's twitterprofil introduceret en ugentligt tilbagevendende gimmick ved navn ”ugens uforståelige” hvor vi præsenterer et sjældent eller mærkeligt ord fra en af vores historiske ordbøger i håb om at gøre brugerne interesserede i selv at gå på opdagelse.

Et eksempel ses på Figur 1; ordet er *brachylogos*, og via klik på linket, sendes brugeren videre til artiklen i Meyers Fremmedordbog. Af besøgsstatistikken fra dagene omkring offentliggørelsen, se Figur 2, fremgår det at vi fik kanaliseret 21 brugere ind på Meyers Fremmedordbog for at se at *brachylogos* betyder ’en, som tilstræber Korthed i Tale’. Det svarede til næsten 5 procent af alle der fulgte DDO's twitterprofil på det tidspunkt, hvilket må siges at være en succes – og viser at vi godt kan lokke brugere ind på de historiske ordbøger, selvom de egentlig har valgt at følge den moderne ordbogs profil.



Figur 1: ”Ugens uforståelige” på DDO's twitterprofil.



Figur 2: Besøgstal på opslaget for *brachylogus* i Meyers Fremmedordbog i dagene omkring promovningen på Twitter.

4. Sammenkædning af ordbøgerne

Vi tror imidlertid at der skal mere til end enkeltstående tiltag – det må handle om at integrere de historiske ordbøger dér hvor brugerne er, så de kan blive en del af den almindelige ordbogsbrug. For ordnet.dk er det derfor afgørende at de historiske ordbøger knyttes tættere til DDO, hvis hjemmeside i foråret 2015 blev besøgt af 70.000 brugere hver dag (ud af 80.000 besøg på hele sitet ordnet.dk).

Vi ønsker således dels at hjælpe brugerne videre fra DDO til de historiske ordbøger hvor det er relevant, dels at øge interessen for aktivt at gå ind i de historiske ordbøger. Det sidste kan f.eks. ske via ”appetitvækkere” som i eksemplet med *brachylogus*.

Løsningen må ligge i en form for sammenkædning af de moderne og de historiske ordbogsressurser, og vi ser her to grundlæggende forskellige strategier – portalløsningen og netløsningen. Ved en portalløsning forstår vi et søgeresultat der viser resultatet fra alle ordbøgerne samtidig, fx som det sker på den danske hjemmeside sproget.dk, der viser opslag fra blandt andet Retskrivningsordbogen, DDO og ODS. Figur 3 viser et eksempel på opslag på sproget.dk.

Vi mener ikke at portalløsningen er en god mulighed for os; dels fordi vores ressourcer er meget forskelligartede – Moths Ordbog er fx skrevet omkring år 1700, og stilen i DDO adskiller sig væsentligt fra de andre – dels fordi et opslag kan komme til at fylde for meget hvis vi viser alle opslag i deres helhed. For større artikler vil det give et rodet og uoverskueligt søgeresultat.

The screenshot shows the website sproget.dk with a search for 'kapre'. The search results are displayed in a central column, with a sidebar on the left containing navigation links like 'FORSIDE', 'NYHEDER', 'RÅD OG REGLER', 'TEMAER', and 'LEG OG LÆR'. The search results section includes a 'Søgeresultat' header, a search bar, and a list of results. The first result is 'Retskrivningsordbogen', followed by 'Den Danske Ordbog' and 'Ordbog over det danske Sprog'. A blue box on the right offers options to refine the search, and a red box at the bottom contains a warning about historical forms in the 'Ordbog over det danske Sprog'.

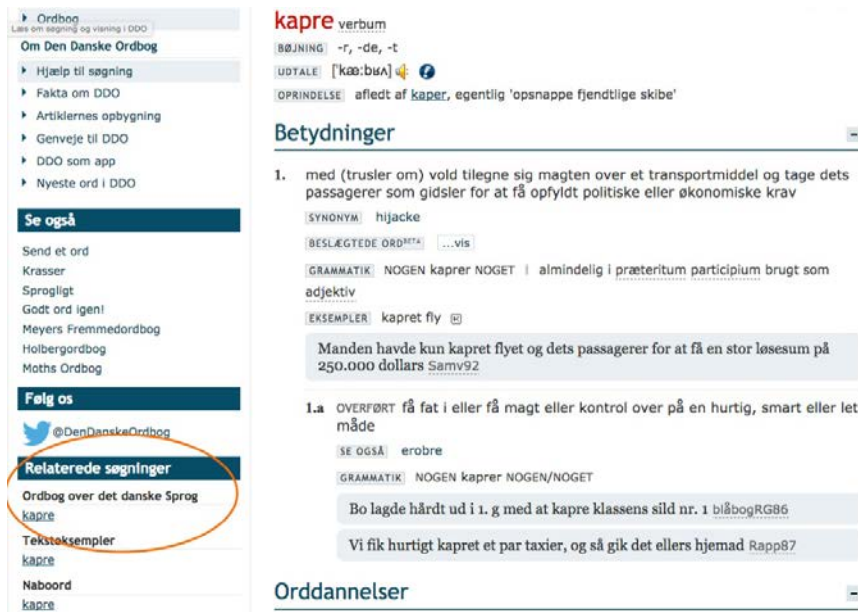
Figur 3: Opslag på portalen sproget.dk, med samlet visning af artikler fra flere ordbøger.

En anden udgave af portalløsningen er kun at vise korte uddrag fra hver enkelt ordbog, men med den model risikerer vi både at brugeren går glip af vigtig information, og at der skal bruges et ekstra klik for at komme til det ønskede resultat.

Vi foretrækker derfor netværkløsningen, hvor tanken er at selvstændige ordbogsressourcer i stedet knyttes sammen ved at henvise til hinanden. På den måde koncentrerer brugeren sig til enhver tid om én ordbog, men tilbydes samtidig muligheden for at skifte til tilsvarende eller beslægtede opslag i de andre ordbøger.

Allerede fra det nuværende ordnet.dk's lancering i 2009 har vi haft en slags netværk mellem DDO og ODS i boksen "Relaterede søgninger", se Figur 4, hvor brugerne kunne skifte til et tilsvarende opslag i den anden ordbog. Problemet er, at boksen i sin nuværende udformning kun virker for de ordbøger der ligger direkte på ordnet.dk-domænet (DDO og ODS), og at

den på hjemmesiden er så diskret placeret, at mange aldrig har opdaget den – og så er det jo så som så med netværket.



Figur 4: Opslag i DDO på ordnet.dk, med link til ODS i venstre kolonne (under "Relaterede søgninger").



Figur 5: Opslag i DDO på ordnet.dk, med link til ODS og Meyers Fremmedordbog i det centrale vindue (under "Andre ordbøger").

Senest har vi derfor introduceret henvisninger på en mere central plads, når en søgning ikke matcher et ord i ordbogen. På Figur 5 vises således resultatet af en søgning på ordet *kemitypi*, som ikke findes i Den Danske Ordbog. I stedet får man henvisninger til ODS og til Meyers Fremmedordbog, hvor ordet optræder.

5. Sammenkædning af ordbogsapper

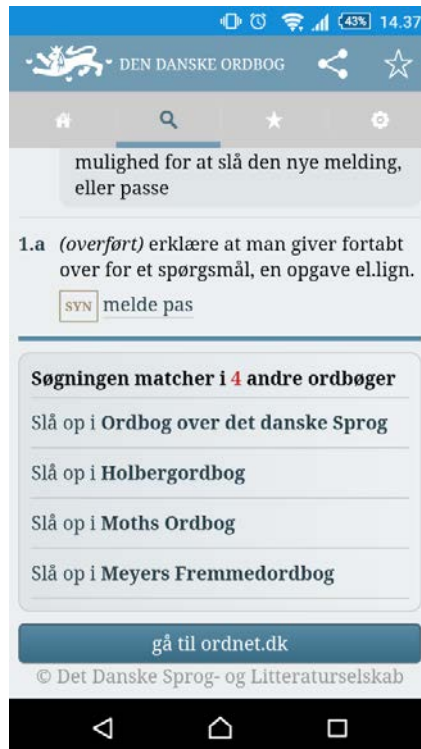
Hvor vi i de traditionelle internetudgaver af ordbøgerne kun tilbyder henvisning til andre ordbøger, hvis ordet ikke findes i den aktuelle ordbog, har vi i vores apper forsøgt at gå et skridt videre med netværkstanken; vi har udgivet selvstændige apper med vores historiske ordbøger, men de er aldrig blevet nær så populære som DDO-appen, se Tabel 2, og i foråret 2015 opdaterede vi derfor vores apper med en funktion der i alle opslag præsenterer henvisningerne til de andre ordbøger hvor der er et match på opslagsordet.

Ordbogsapp	Downloads (Android og iOS)
DDO	130.000
ODS	17.000
Moths ordbog	3.200
Holbergordbog	3.000
Meyers Fremmedordbog	1.500

Tabel 2: Samlet antal downloads af DSL's ordbogsapper medio maj 2015.

Figur 6 viser således et opslag på ordet *pas* op i DDO-appen, med henvisninger til de fire historiske ordbøger i bunden af artiklen. Henvisningerne fører brugeren direkte over til den relevante artikel, på tværs af apperne, og hvis man ikke har installeret mål-appen, får man mulighed for at installere den via et klik i et pop op-vindue.

Der er således tale om et relativt simpelt forsøg på at implementere netværkstanken – som en afprøvning af om vi på den måde kan lokke vores brugere til at gå på opdagelse i de historiske ordbøger.



Figur 6: Henvisninger fra DDO-appen til de historiske ordbogsapper.

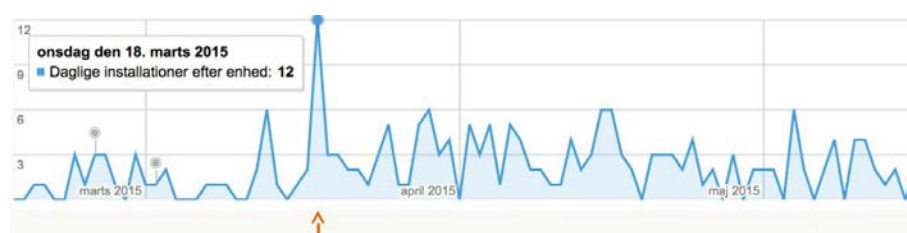
Vi ville jo gerne nu kunne konkludere at dette tiltag har virket, men som det ses af downloadstatistikkerne, Figur 7-9, kan vi kun påvise en relativt lille effekt. Figur 7 viser downloadstatistikken for appen for Moths Ordbog på Android-plattformen. Pilen markerer den dag funktionen blev offentliggjort, og der ses en lille stigning i app-hentninger fra 2-4 til 4-8 om dagen. Det er en fordobling, men det kan næppe anses for at være signifikant, når man tænker på hvor små de absolutte tal er.

Det samme billede gør sig gældende for Holbergordbogen, se Figur 8. Antallet af downloads er steget fra 1-3 til 2-6 om dagen. Igen en slags fordobling, men igen meget små tal.

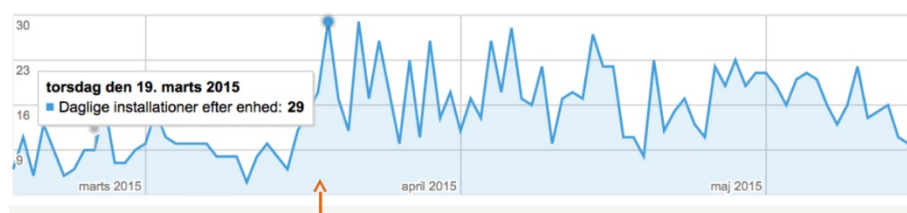
Af de historiske ordbogsapper er ODS den mest populære, med cirka 17.000 downloads i alt. Også her er billedet det samme, se Figur 9, med en fordobling af det daglige antal downloads til i gennemsnit 20 om dagen på Android-plattformen.



Figur 7: Antal downloads af appen med Moths Ordbog i dagene omkring offentliggørelse af funktionen med linkning mellem ordbogsapperne.



Figur 8: Antal downloads af appen med Holbergordbogen i dagene omkring offentliggørelse af funktionen med linkning mellem ordbogsapperne.



Figur 9: Antal downloads af appen med ODS i dagene omkring offentliggørelse af funktionen med linkning mellem ordbogsapperne.

Disse tal skal ses i sammenhæng med at vi registrerer cirka 10.000 søgninger i vores apper om dagen, så vi må konkludere at netværksløsningen – i hvert fald i sin nuværende form – ikke for alvor har formået at skabe et boost til de historiske ordbøger.

Men vi har ikke givet op endnu. Indtil nu har vores netværk været ”dumt” – der er tale om simple søgninger og henvisninger til de andre ordbøger, og vi planlægger derfor at gå et skridt videre og for alvor knytte ordbøgerne sammen. Med en dybere sammenkædning kan vi – forhåbentlig – skabe et netværk der er endnu mere relevant for brugerne.

6. Det fortsatte arbejde

Vi arbejder henimod at etablere en central database der sammenkæder alle DSL's ordbogsressurser. Det første skridt er at fastlægge strukturen for en sådan database. Vi håber på derefter at få mulighed for at begynde at bruge den i større skala, dvs. i første omgang ved at inddrage lemmaer fra DDO og ODS, sidenhen ved at tilføje nye og ikke mindst gamle ord. En oplagt synergi-mulighed vil her være ordforrådet i Gammeldansk Ordbog (GO), idet der for tiden i forbindelse med projektet ”Middelalderens danske sprog og litteratur formidlet til nutiden” bl.a. arbejdes på dels at lemmatisere ordene i en række middelaldertekster (ved linkning til GO-lemmaer) på hjemmesiden middelaldertekster.dk, dels at linke GO-lemmaerne til deres moderne form, hvis der findes en sådan.

En sådan sammenkædning vil dels kunne give sikre links til samme opslagsord i andre ordbøger, dels kunne bidrage til forbedring af ordbøgernes søgefunktion, idet ikke bare den moderne stavemåde, men alle perioders stavemåder vil kunne inddrages ved opbygningen af søgealgoritmer. Således kan en bruger der i dag ønsker at slå *afreth* op i GO, kun finde ordet, hvis han/hun kender den historiske (og normaliserede stavemåde); med en lemma-sammenkædning vil ordet også kunne findes ved opslag på den moderne form, *ævred*, og man vil desuden kunne henvise til de relevante opslag i alle de moderne og historiske ordbøger.

Desuden vil man naturligvis kunne inddrage flere ordbogskilder, f.eks. Otto Kalkars ”Ordbog til det ældre danske Sprog”, der for tiden er ved at blive digitaliseret, men også former fra f.eks. korpustekster kunne tænkes at indgå, ligesom uofficielle varianter og stavfejl kunne inddrages og bidrage til forbedring af søgemulighederne.

Litteratur

Ordbøger

DDO (2003) = *Den Danske Ordbog*. Hovedredaktører: Ebba Hjort & Kjeld Kristensen. København: Gyldendal. <ordnet.dk/ddo> (november 2015).

- GO (2015) = *Gammeldansk Ordbog*. Ledende redaktør: Marita Akhøj Nielsen. <gammeldanskordbog.dk> (december 2015, beta).
- Holbergordbog = Aage Hansen m.fl. (1981-88): *Holberg-Ordbog*. Ordbog over Ludvig Holbergs Sprog I-V. København: Det Danske Sprog- og Litteraturselskab. <holbergordbog.dk> (november 2015).
- Kalkar = Kalkar, Otto (1881-1918): *Ordbog til det ældre danske Sprog. I-V*. København: Akademisk Forlag.
- Meyers Fremmedordbog = Meyer, Ludvig (1924): *Fremmedordbog*. København: Gyldendalske Boghandel. <meyersfremmedordbog.dk> (november 2015).
- middelaldertekster.dk = *Middelalderens danske sprog og litteratur formidlet til nutiden*. Ledende redaktør: Marita Akhøj. København: Det Danske Sprog- og Litteraturselskab. <middelaldertekster.dk> (december 2015).
- Moths Ordbog (2013) = Marita Akhøj Nielsen m.fl. (red.): *Moths Ordbog*. København: Det Danske Sprog- og Litteraturselskab. <mothsordbog.dk> (november 2015).
- ODS (1918-55) = Dahlerup, Verner m.fl.: *Ordbog over det danske Sprog*. København: Det Danske Sprog- og Litteraturselskab. <ordnet.dk/ods> (november 2015).
- ODS-S = Anne Duekilde m.fl. (1992-2005): *Ordbog over det danske Sprog. Supplement*. København: Gyldendal. <ordnet.dk/ods> (november 2015).

Anden litteratur

- Eegholm-Pedersen, Sv. (2013): Ministerens ordbog. Matthias Moths store danske ordbog 1700-1717. I: *Magasin fra Det kongelige Bibliotek*, 26(4), København: Det Kongelige Bibliotek, 23-37.

Thomas Troelsgård
seniorredaktør, cand.mag.
Det Danske Sprog- og Litteraturselskab
Christians Brygge 1
DK-1219 København K
tt@dsl.dk

Nicolai Hartvig Sørensen
seniorredaktør, cand.mag.
Det Danske Sprog- og Litteraturselskab
Christians Brygge 1
DK-1219 København K
nhs@dsl.dk

Stamtræer og digitale etymologiske ordbøger

Seán D Vrieland

Presentation of etymological information in dictionaries varies widely, making it difficult to propose a standard form of encoding this information in the process of digitization. The following proposes an extension of the TEI P5 standard for encoding texts using XML, wherein a new element <etymon> is able to nest recursively, thereby allowing the XML format to match that of a language family tree. Problems of reconstructed proto-forms, loanwords, and language grouping are further discussed.

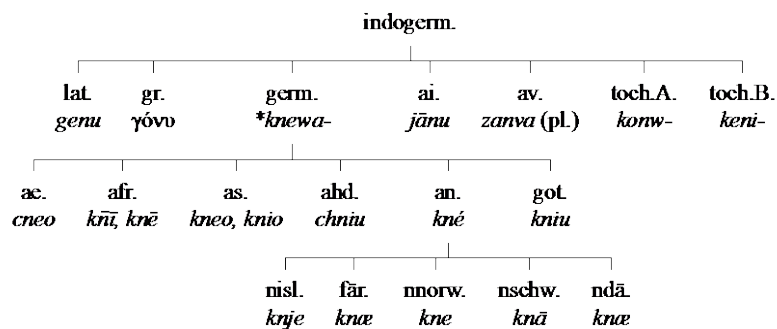
1. Digitale ordbøger og etymologier

Når ordbøger skal digitaliseres, opstår der problemer med at kunne opmærke etymologiske oplysninger på en måde, så de kan læses både af brugeren og computeren. Brugeren vil for eksempel opfatte kognaterne for oldnordisk *kné* i de Vries' (1961) etymologiske ordbog på en helt anden måde, end computeren gør.

kné n. 'knie; glied; krummholz' (< germ. **knewa-*), nisl. *knje*, fär. *knæ*, nnorw. *kne*, nschw. *knä*, ndä. *knæ*. – got. *kniu*, æ. *cneo*, afr. *knī*, *knē*, as. *kneo*, *knio*, ahd. *chniu*. – lat. *genu*, gr. γόνυ, ai. *jānu*, av. *zanva* (pl.), toch.A. *kanw-*, B. *keni-* 'knie' (s. über das verhältnis der formen Petersson SVS, Lund, 1, 1921, 7).

For computeren er det muligt at opmærke alle kognaterne som beslægtede med *kné*, men brugeren forstår disse informationer anderledes. Der er stor forskel på den fællesgermanske form **knewa-*, som er en rekonstrueret urform – den oldnordiske forms forfader – og dansk *knæ*, der er en af den oldnordiske forms efterkommere.

Brugeren vil også opfatte tankestregerne i de Vries' ordbog som adskillellestegn. De nordiske former, der står foran den første tankestreg, og som er opslagsordets efterkommere, er beslægtet med *kné* på en anden måde end de efterfølgende former. Disse former er endvidere placeret i to grupper: germanske kognater, som er tæt beslægtede med *kné*, og ikke-germanske, indoeuropæiske former, som også er beslægtede, men længere tilbage i tiden.



Figur 1: Oldnordisk *kné* i et stamtræ.

En bruger, der har forstået den måde, de Vries (1961) viser de etymologiske informationer på, vil anbringe *kné* og dets kognater i et slags stamtræ som i Figur 1. Spørgsmålet er altså, hvordan disse kognater kan opmærkes med præcis samme struktur.

1.1. Tidligere løsninger til opmærkning af etymologier

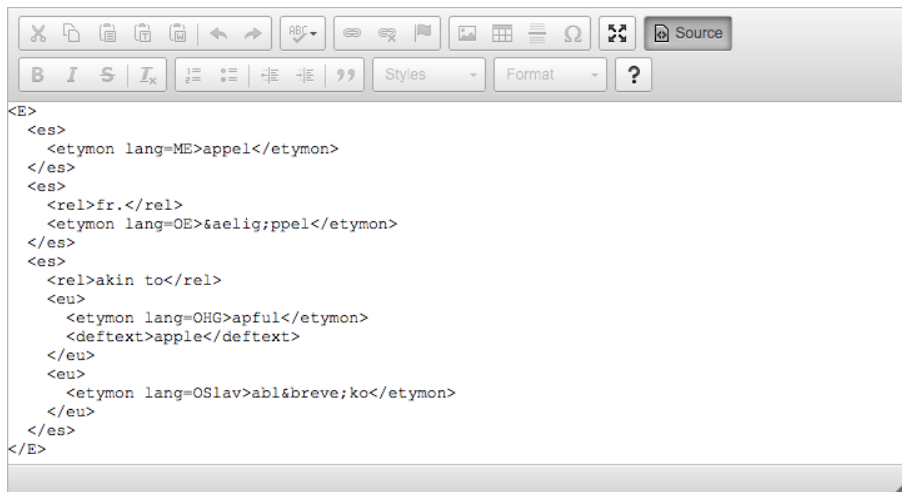
Sean Crist (2005) diskuterer teoretisk, hvordan etymologiske informationer kunne opmærkes. Han skelner mellem tre opmærkningstyper:

Type I. Markup schemes which make no provision for etymological data.

Type II. Markup schemes where etymological data is delimited as such, but treated as unstructured prose

Type III. Markup schemes where the mathematical relationships recognized in historical/comparative linguistics are somehow embodied in the markup system in (semi-)machine-readable form

Som eksempel på type III diskuterer Crist et forslag fra Amsler & Tompa (1988), der betjener sig af et antal SGML-tags til opmærkning af ordbøger. Etymologier i Amsler & Tompas system grupperes i et tag <E> (*etymon*), der indeholder tagget <es> (*etymological segment*) som ”branches of a universal etymology tree” (Amsler & Tompa 1988:7) – dvs. ordets slægtskab med en kognat – og <eu> (*etymological unit*) til kognaten og dens betydning. Hver kognat tagges i <etymon> og forsynes med sprogets navn som attribut (for eksempel <etymon lang=OHG> for den oldhøjtske kognat), og informationer om, hvordan kognaterne er beslægtede, tagges i <rel> (*relation*). Et eksempel fra Amsler & Tompa på etymologien af engelsk *apple* er gengivet i Figur 2.



```

<E>
  <es>
    <etymon lang=ME>appel</etymon>
  </es>
  <es>
    <rel>fr.</rel>
    <etymon lang=OE>æaelig;ppel</etymon>
  </es>
  <es>
    <rel>akin to</rel>
    <eu>
      <etymon lang=OHG>apful</etymon>
      <deftext>apple</deftext>
    </eu>
    <eu>
      <etymon lang=OSlav>ablǫbreve;ko</etymon>
    </eu>
  </es>
</E>

```

Figur 2: Etymologien til engelsk *apple* ifølge Amsler & Tompa (1988:4).

Amsler & Tompas forslag giver mulighed for at opmærke kognaterne samt deres definitioner og nogle informationer om, hvordan de er beslægtet med opslagsordet. Men computeren ville ikke kunne læse denne kode og vise kognaterne som et stamtræ, hvor det moderne engelske ord kommer fra det middelengelske, der selv kommer fra det oldengelske. Der gives heller ikke nogen informationer om, hvordan de oldhøjtyske og oldslaviske former er beslægtet med hinanden, herunder ikke mindst hvordan de er beslægtet med opslagsordet.

1.2. Digitalisering af ordbøger ifølge TEI

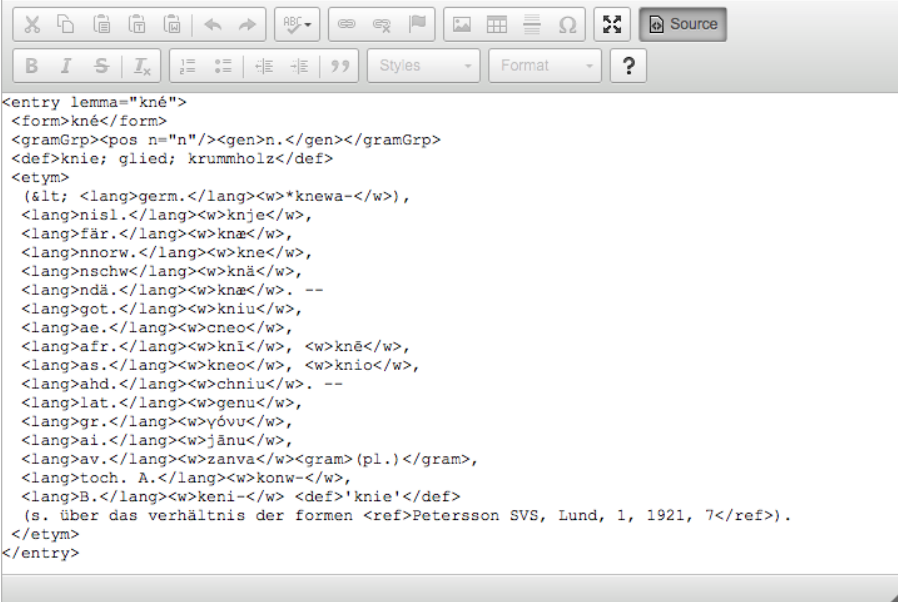
Konsortiet TEI (*Text Encoding Initiative*) blev grundlagt i 1987 med et særligt formål: at udvikle retningslinjer til opmærkning af tekster. Deres første version af *Guidelines for Electronic Text Encoding and Interchange* udkom i 1994.¹

TEIs første *Guidelines* beskrev, hvordan forskellige typer tekster kunne opmærkes i en SGML-standard, blandt andet ordbøger. Den nuværende version – P5 – er opdateret fra SGML til XML og inkluderer flere tags til opmærkningen af ordbøger, bl.a. tagget <etym>, som indeholder alle etymologiske oplysninger og kan sammenlignes med Amsler & Tompas <E>-tag. Men når det handler om etymologier, siger TEI P5 tydeligt:

Etymologies may contain highly structured lists of words in an order indicating their descent from each other, but often also include related words and forms outside the direct line of descent, for comparison. Not infrequently, etymologies include commentary of various sorts, and can grow into short (or long!) essays with prose-like structure. This variation in structure makes it impracticable to define tags which capture the entire intellectual structure of the etymology or record the precise interrelation of all the words mentioned. (TEI P5:9.3.4 "Etymological Information").

¹ Disse retningslinjer beskriver den tredje version af TEI's opmærkningsstandard, dvs. P3. Se endvidere Ide & Sperberg-McQueen (1995).

Som TEIs retningslinjer ser ud nu, er det muligt at opmærke for eksempel *kné* fra de Vries (1961) som i Figur 3. En af grundene til at anvende TEI P5 er, at mange informationer kan opmærkes således, at den trykte ordbogs udseende kan bibeholdes. Det er for eksempel muligt at angive ordet *kné* som substantiv i tagget <pos> (*part of speech*), selvom dette ikke vises i de Vries' trykte version. Men de etymologiske informationer kan kun opmærkes ustruktureret; det er ikke muligt at gruppere de nordiske former som efterkommere eller de germanske kognater som tættere beslægtet.



```

<entry lemma="kné">
  <form>kné</form>
  <gramGrp><pos n="n"/><gen>n.</gen></gramGrp>
  <def>knie; glied; krummholz</def>
  <etym>
    (&lt; <lang>germ.</lang><w>*knewa-</w>),
    <lang>nisl.</lang><w>knje</w>,
    <lang>fär.</lang><w>knæ</w>,
    <lang>nnorw.</lang><w>kne</w>,
    <lang>nschw.</lang><w>knä</w>,
    <lang>ndä.</lang><w>knæ</w>. --
    <lang>got.</lang><w>kniu</w>,
    <lang>ae.</lang><w>cneo</w>,
    <lang>afr.</lang><w>kni</w>, <w>kné</w>,
    <lang>as.</lang><w>kneo</w>, <w>knio</w>,
    <lang>ahd.</lang><w>chniu</w>. --
    <lang>lat.</lang><w>genu</w>,
    <lang>gr.</lang><w>γόνυ</w>,
    <lang>ai.</lang><w>jānu</w>,
    <lang>av.</lang><w>zanva</w><gram>(pl.)</gram>,
    <lang>toch. A.</lang><w>konw-</w>,
    <lang>B.</lang><w>keni-</w> <def>'knie'</def>
    (s. über das verhältnis der formen <ref>Petersson SVS, Lund, 1, 1921, 7</ref>).
  </etym>
</entry>

```

Figur 3: Opslagsordet *kné* ifølge TEI P5.

2. Et nyt forslag

En fordel ved at bruge XML til opmærkning af etymologiske informationer er muligheden for *nesting*, dvs. at et tag kan indeholde et andet tag af samme type. Som løsning på problemet med stamtræer og digitale etymologiske ordbøger vil jeg foreslå et nyt tag, der kan indgå i TEIs retningslinjer. Tagget <etymon>, der bør placeres inde i TEIs <etym>-tag, vil kunne nestes rekursivt og indeholde tags med en kognats informationer, herunder om dets efterkommere.

Hvis vi for eksempel vil opmærke etymologien af engelsk *apple*, kan vi danne et tag <etymon>, der indeholder <w>apple</w>. Dette <etymon>-tag placeres i et andet <etymon>-tag, der indeholder den middelengelske form <w>appel</w>. Endelig anbringes dette <etymon>-tag i et tredje tag, nemlig det, der indeholder den oldengelske form <w>æppel</w>. Informationer om sprogene kan opmærkes som attributter i <etymon>-tagget, dvs. <etymon xml:lang=ang> til den oldengelske form.

Det vil også være vigtigt at kunne angive forskelle mellem former, der er belagt i skriftsproget, for eksempel oldnordisk *kné*, og rekonstruerede urformer som fællesgermansk **knewa-*. Dette kan gøres ved hjælp af et nyt attribut til <etymon>-tagget, nemlig *@attested*. På den måde vil **knewa-* kunne opmærkes i <etymon xml:lang=gem attested=no>.

Problemet med at gruppere for eksempel alle indoeuropæiske kognater kan løses ved hjælp af et tag <etymon xml:lang=ine attested=no>, som ikke indeholder et <w>-tag, fordi der ikke er angivet nogen urindoeuropæisk form hos de Vries.

Figur 4 viser, hvordan alle kognaterne placeres i indlejrede <etymon>-tags. Det yderste tag står for indoeuropæisk, som indeholder samtlige kognater, fordi de alle er indoeuropæiske sprog. Det næste tag er germansk, som indeholder den urgermanske form og dens efterkommere, de germanske kognater. Dernæst kommer de nordiske sprog, som er placeret i et <etymon>-tag, der står for oldnordisk. Til slut kommer de øvrige germanske sprog (inden for det germanske <etymon>-tag) og indoeuropæiske sprog (inden for det indoeuropæiske <etymon>-tag).


```

<entry lemma="kné">
  <form>kné</form>
  <gramGrp><pos n="n"/>
  <gen>n.</gen></gramGrp>
  <def>knie; glied; krummholz</def>
  <etym>
    <etymon xml:lang="ine" attested="no">
      <etymon xml:lang="gem" attested="no">&lt; <lang>germ.</lang><w>knewa-</w>,</etymon>
      <etymon xml:lang="non">
        <etymon xml:lang="is"><lang>nisl.</lang><w>knje</w>,</etymon>
        <etymon xml:lang="fa"><lang>fär.</lang><w>knæ</w>,</etymon>
        <etymon xml:lang="nn"><lang>nnorw.</lang><w>kne</w>,</etymon>
        <etymon xml:lang="sv"><lang>nschw.</lang><w>knä</w>,</etymon>
        <etymon xml:lang="da"><lang>ndä.</lang><w>knæ</w>.</etymon>
      </etymon>
      <etymon xml:lang="got"> -- <lang>got.</lang><w>kniu</w>,</etymon>
      <etymon xml:lang="ang"><lang>ae.</lang><w>cneo</w>,</etymon>
      <etymon xml:lang="ofs"><lang>afr.</lang><w>kni</w>,<w>kné</w>,</etymon>
      <etymon xml:lang="osx"><lang>as.</lang><w>kneo</w>,<w>knic</w>,</etymon>
      <etymon xml:lang="goh"><lang>ahd.</lang><w>chniu</w>.</etymon>
    </etymon>
    <etymon xml:lang="lat">-- <lang>lat.</lang><w>genu</w>,</etymon>
    <etymon xml:lang="grc"><lang>gr.</lang><w>γόυυ</w>,</etymon>
    <etymon xml:lang="sa"><lang>ai.</lang><w>jānu</w>,</etymon>
    <etymon xml:lang="ae"><lang>av.</lang><w>zanva</w><gram>(pl.)</gram>,</etymon>
    <etymon xml:lang="xto"><lang>toch. A.</lang><w>konw-</w>,</etymon>
    <etymon xml:lang="txb"><lang>B.</lang><w>keni-</w></etymon><def>'knie'</def>
  </etymon>
  (s. über das verhältnis der formen <ref>Petersson SVS, Lund, 1, 1921, 7</ref>).
</etym>
</entry>

```

Figur 4: Opslagsordet *kné* i TEI samt etymologiske informationer.

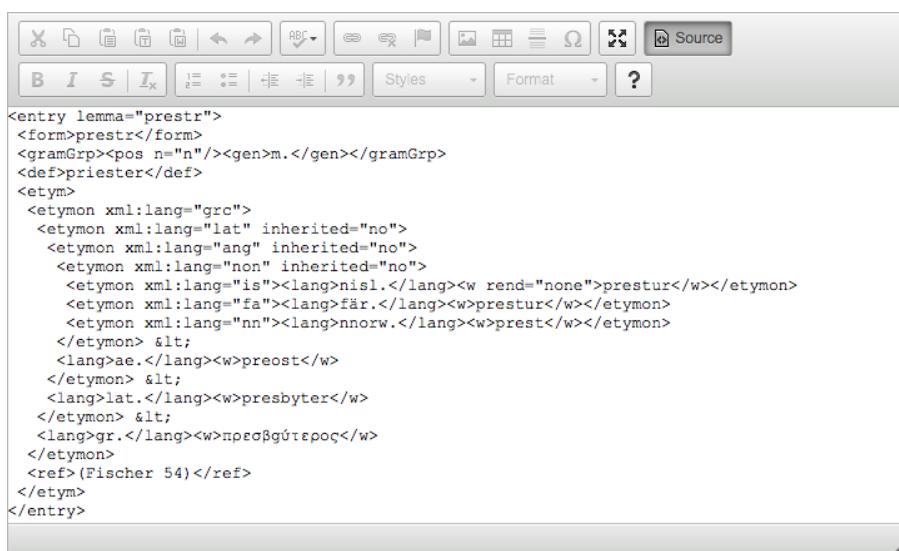
2.1. Arveord og låneord

I eksemplet *kné* er det muligt at bruge tagget `<etymon>` til opmærkning af etymologien, ikke mindst fordi alle sprogene er beslægtet og alle kognater nedarvet. Etymologiske ordbøger indeholder ikke kun nedarvede ord, men også et antal låneord, for eksempel *prestr* i de Vries (1961).

prestr m. 'priester', nisl. fär. *prestur*, nnorw. *prest* < ae. *preost* < lat. *presbyter* < gr. πρεσβύτερος 'älter' (Fischer 54).

Ifølge Crist (2005) er der en afgørende forskel mellem arveord og låneord, som bør opmærkes på to helt forskellige måder. Vi kunne forestille os endnu et nyt tag `<loan>`, men det ville ikke gøre det nemmere for computeren. Men i koden behøver der ikke være så stor forskel på arveord og låneord, især hvis alle `<etymon>`-tags har sprogenes navne som attributter. I stedet foreslår jeg et andet attribut *@inherited*, som angiver, om et `<etymon>` er et arveord eller et låneord. For eksempel kunne oldnordisk *prestr* placeres i tagget `<etymon xml:lang=non attested=yes inherited=no>`.

Opmærkningen af *prestr* vises i Figur 5. Bemærk også, at de islandske og færøske former er identiske og derfor kun anføres én gang med den samme form (*prestur*). For at kunne inkludere den islandske form opmærkes den i et `<w>`-tag med attributtet *@rend=none*, dvs. at det ikke vises i den trykte bog.



```

<entry lemma="prestr">
  <form>prestr</form>
  <gramGrp><pos n="n"/><gen>m.</gen></gramGrp>
  <def>priester</def>
  <etym>
    <etymon xml:lang="grc">
      <etymon xml:lang="lat" inherited="no">
        <etymon xml:lang="ang" inherited="no">
          <etymon xml:lang="non" inherited="no">
            <etymon xml:lang="is"><lang>nisl.</lang><w rend="none">prestur</w></etymon>
            <etymon xml:lang="fa"><lang>fär.</lang><w>prestur</w></etymon>
            <etymon xml:lang="nn"><lang>nnorw.</lang><w>prest</w></etymon>
          </etymon> &lt;
          <lang>ae.</lang><w>preost</w>
        </etymon> &lt;
          <lang>lat.</lang><w>presbyter</w>
        </etymon> &lt;
          <lang>gr.</lang><w>πρεσβύτερος</w>
        </etymon>
        <ref>(Fischer 54)</ref>
      </etym>
    </etym>
  </entry>

```

Figur 5: Opslagsordet *prestr* i TEI samt etymologiske informationer.

3. Konklusion

Brugen af XML-opmærkning af etymologiske ordbøger efter TEI's retningslinjer giver mange muligheder for brugeren. Som tidligere nævnt kan man dermed opmærke informationer om både tekstens udseende og tilføje ekstra oplysninger, men der bliver tillige skabt mulighed for, at en opmærket ordbog kan bruges sammen med andre TEI-opmærkede tekster i for eksempel digitale udgaver.

Tekster behøver ikke være på papir, før de opmærkes i TEI; retningslinjerne kan også bruges til digitalt udviklede tekster. Dette er en af de vigtigste grunde til, at jeg ville udvide TEI's retningslinjer med tagget <etymon>. Som en del af mit ph.d.-projekt anvender jeg TEI P5 til opmærkning af middelalderlige håndskrifter på et bestemt oldnordisk sprog, nemlig oldgutnisk. Når jeg er færdig med at opmærke et håndskrift, hvor alle ord er tagget med et lemma, er det muligt at eksportere alle lemmata til en ekstern fil og derfra bygge en ordbog over det oldgutniske sprog. Ved anvendelse af <etymon>-tagget vil det være muligt at angive hvert ords kognater på de andre oldnordiske sprog samt en etymologi med andre germanske eller indoeuropæiske kognater.

Man kan forvente flere problemer ved anvendelse af dette tag, end jeg har beskrevet i denne artikel. For eksempel står det ikke helt klart, hvordan man kan opmærke sammensatte ord, hvor hvert led har sin egen etymologi. Men brug af et tag, som indeholder et sprogs informationer, og som kan anbringes inden i hinanden rekursivt med henblik på at beholde et stamtræs struktur, vil skabe mange nye muligheder for brugere af digitale etymologiske ordbøger.

Litteratur

Ordbøger

de Vries, Jan (1961): *Altnordisches etymologisches Wörterbuch*. Leiden: Brill.

Anden litteratur

- Amsler, Robert A. & Frank Wm. Tompa (1998): An SGML-based Standard for English Monolingual Dictionaries. I: *Information in Text: Fourth Annual Conference of the UW Centre for the New Oxford English Dictionary: Proceedings of the Conference, October 26-28, 1988, Waterloo, Canada*. Waterloo, Canada, 61-79.
- Crist, Sean (2005): Toward a formal markup standard for etymological data. Oakland, CA: Linguistic Society of America Annual Meeting. <www.sean-crist.com/professional/publications/crist_etym_markup.pdf> (august 2015).
- Ide, Nancy M. & C. M. Sperberg-McQueen (1995): The TEI: History, Goals, and Future. I: *Text Encoding Initiative: Background and Context*. Dordrecht: Kluwer Academic Publishers.
- TEI P5 = TEI Consortium (red.) (2015): *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Version 2.8.0. TEI Consortium. <www.tei-c.org/Guidelines/P5/> (august 2015).

Seán D Vrieland
ph.d.-stipendiat
Nordisk Forskningsinstitut
Københavns Universitet
Njalsgade 136
DK-2300 København S
sean.vrieland@gmail.com

GENERELT PROGRAM

Konventionelle figurative enheder (idiomer) i Jysk Ordbog og andre større ordbøger

Torben Arboe

Inside the framework of cognitive linguistics and lexical semantics the study employs a theory of conventional figurative units, e.g. metaphors and similes, based on a conceptual theory of metaphors. The lexical material is found in the card index and files of the *Jysk Ordbog (Dictionary of the Jutland Dialects)*; it consists of idioms from a broad range of semantic fields, arranged in oppositions as 'richness vs. poorness', 'cleverness vs. foolishness', and it covers especially topic groups of the daily work and life on farms, in villages etc. of the countryside. The study shows how the idioms are treated in the entries of *Jysk Ordbog*, and special attention is given to instances of polysemy due to figurative use of word; this may require addition of a submeaning to a lexical entry. The results are summarized in an overview chart for the key lexemes of the idioms in the respective semantic fields.

Som led i mine undersøgelser af, hvordan man i ordbøger mere systematisk behandler fraseologismer, dvs. faste ordforbindelser af forskellig art (jf. Arboe 2008, 2010), vil jeg her se nærmere på faste forbindelser med overførte betydninger, idiommer mv., først og fremmest metaforer og analogier.

1. Teoretisk baggrund: kognitiv lingvistik, figurativ metafor-teori og leksikalsk semantik

Traditionelt definerer man i retorikken og litteraturteorien en metafor som et udtryk med en overført eller figurativ betydning, mens man kun tilskriver analogier en ligefrem betydning. Sætningen *Hun er en rose* kan kun være en metafor, fordi en person ikke kan være en blomst, og fordi en rose ofte omtales som en smuk blomst, i hvert fald i den europæiske kulturkreds: Her har subst. *rose* en symbolsk funktion, ordet kan klassificeres som et kulturelt symbol (Dobrovolskij & Piirainen 2005:35). Sætningen *Hun er som en rose* vil man derimod traditionelt ikke se som en metafor, men som en analogi,

også kaldet simile. Hvis man derimod anlægger en bredere lingvistisk/pragmatisk synsvinkel, er afstanden mellem metafor og analogi ikke så stor. Dette påpeges inden for den kognitive lingvistik af bl.a. Croft & Cruse 2004:211ff., og inden for den kognitive metafor-teori ud fra især Lakoff and Johnson (1980/2003) og (Lakoff 1987) bruger man hellere blot betegnelsen *idiomer* om bl.a. de faste forbindelser, der tidligere traditionelt blev kaldt metaforer (Dobrovolskij & Piirainen 2005:92). Det kan bemærkes, at hvis man vil bruge en analogi, vælger man naturligvis det bedst mulige eksempel, man kan finde; hvis andre finder dette tilstrækkelig prægnant, bruger de det måske også, hvorved det efterhånden kan blive en fast ordforbindelse næsten på linje med en metafor (som i parentes bemærket må være opstået og udbredt på lignende måde). Herudfra foretrækker jeg termen *figurativ enhed* for både metaforer og analogier, og da Jysk Ordbog og andre ordbøger registrerer talt og eller skrevet sprog i almindelighed, ikke specielt digterisk sprog el.lign., kan man bruge betegnelsen *konventionel figurativ enhed* om disse typer faste ordforbindelser.

2. Leksikalsk materiale

Mit materiale består af en lang række idiomere i jysk inden for det centrale ordforråd vedr. dagliglivet, idet man i dette ordforråd må forvente at finde de fleste og fastest etablerede metaforiske ordforbindelser og talemåder. Disse idiomere er især fundet via excerpering af længere afsnit af Feilberg (1886-1914), desuden ved gennemgang af kartotekssedlerne i Jysk Ordbog for mange enkeltords vedkommende. De opslagsord (leksemer), idiomerne er knyttet til, kaldes *basisleksemer* i det følgende. Ved nogle idiomere kan der opstå tvivl om, hvilket opslagsord (af to eller måske tre) idiomet er tættest knyttet til; dette diskuteres i de enkelte tilfælde, og der henvises fra det valgte basisleksemer til de øvrige.

2.1. Semantiske felter: antonyme hhv. komplementære begreber

Idiomerne er inddelt i to grupper: de, hvis betydning kan beskrives ud fra antonyme adjektiver over for de, der kan beskrives ud fra komplementære adjektiver, som skitseret i Figur 1.

Figur 1: **Begrebsmodsatninger**. Tolinjede skalaer ('biscalar'; Croft and Cruse 2004:167-181).

- 1) ANTONYME: '----- rig ----->
<---- fattig -----'
- [skalaerne er *parallel equipollent*, dvs. overlappende, af samme styrke]
- DELVIS ANTONYME: '----- god ----->
<--- dårlig ----' (svarer til en del af 'god'-skalaen).
- 2) KOMPLEMENTÆRE: <---- kold ----' [0°C] '---- varm ---->
[skalaerne er *disjunct equipollent*, dvs. adskilte, af samme styrke¹]

Til disse typer adjektiver svarer substantiverne 'rigdom vs. fattigdom', 'kulde vs. varme' m.fl., som dermed kan siges at betegne antonyme hhv. komplementære begreber; adjektiverne repræsenterer disse begreber, fx er 'rig' synonymt med 'det at have rigdom'. Inden for antonyme begrebspar findes især idiomere på skalaerne 'klog vs. dum', 'langsom vs. hurtig', 'ødsel vs. nærig', med de tilsvarende substantiviske begreber 'klogskab vs. dumhed' etc.; inden for de komplementære tilsvarende især på skalaen 'heldig vs. uheldig', 'munter vs. gnaven', 'selvisk vs. samarbejdende' og 'levende vs. død', med de tilsvarende substantiviske begrebspar 'held vs. uheld' etc. I det følgende anvendes oftest de substantiviske betegnelser. De foretagne opdelinger viser sig praktiske derved, at fx 'rigdom' og 'fattigdom' ikke placeres på absolutte, adskilte skalaer, men sådan, at fx *mindre rig* modsvares af *mere fattig*, jf. pilenes retning på de to skalaer².

2.2. Opdeling i sagområder

De basisleksemer (opslagsord), som indgår i undersøgelsen, findes inden for især følgende 8 sagområder: a) gården, redskaber, tilbehør mv., b) værksted, værktøj, c) dyr, både husdyr og dyr fra omgivelserne, d) køkken: redskaber, produkter mv., e) inventar i stuehus, f) beklædning mv., g) penge, h) geo-

¹ Eksemplet her fra Croft & Cruse (2004:170) må dog siges kun på et praktisk, dagligdags niveau at vise skalaer af samme styrke: i princippet kan 'varm' omfatte numerisk meget høje temperaturer (millioner af grader celcius), mens 'kold' derimod kun kan omfatte temperaturer ned til minus 273 grader celcius, temperaturernes absolutte nulpunkt.

² Måske kan to af de komplementære begrebspar ('held vs. uheld', 'munterhed, godt humør vs. gnavenhed, vrede') også betragtes som delvis antonyme.

grafi. Der er tale om en praktisk opdeling ud fra de idiommer (eller konventionelle figurative enheder), jeg fandt under excerperingen af Feilberg mv.; men den repræsenterer også adskillige centrale områder fra andre, mere detaljerede, saglige opdelinger, så som Skautrup (1927-1979, jf. oversigten bd. II:241-243) og *Ømålsordbogen* (1992:5). Den anvendte sagopdeling lægger op til, at der som basisleksemer især vælges substantiver (betegnelser for genstande mv.). De behandlede basisleksemer og deres tilhørende sagområder og semantiske felter vises til sidst i et oversigtsskema (afsnit 4).

3. Undersøgelse inden for de respektive semantiske felter

Her behandles først antonyme begrebspar (afsnit 3.1-3.4), dernæst komplementære (afsnit 3.5-3.8). Idiommerne er nummereret fortløbende for at sikre præcise henvisninger fra teksten. Hvis et givet idiom ligger inden for den redigerede del af Jysk Ordbog (A-KL), nævnes pågældende artikel og eventuelle underbetydning. Hvis et idiom har givet anledning (eller været medvirkende) til polysemi i en leksikalsk enhed, nævnes det ved pågældende underbetydning.

3.1. RIGDOM vs. FATTIGDOM

Inden for dette semantiske felt har jeg indtil videre fundet 6 metaforer eller idiommer, hvor nogle også kendes i rigsdansk. Vedr. RIGDOM drejer det sig om: (1) *være ved skillingen* og (2) *have noget på kistebunden*, begge = 'være rig'; desuden (3) *hans plov går rigtigt* 'han tjener mange penge', hvor der tages udgangspunkt i markarbejdet med ploven. De tre sidste gælder FATTIGDOM: i (4) *han gik baglæns af træskoene* = 'han gik bankerot, fallit' er underforstået, at det at *gå baglæns* overført betegner tilbagegang generelt; talemåden med *af træskoene* er ikke nævnt i Jysk Ordbog under *baglæns*, men medtages antagelig ved subst. *træsko*. I (5) *han skræj å æ skjøssel* (dvs. skred af skydselen, vestjysk) = 'gik bankerot, fallit', er udgangspunktet bagning i stor, muret ovn, hvor bageredskabet en skydsel ('bagespade' el.lign.) bruges til at sætte de enkelte brød ind på pladen i ovnen med. Hvis et brød falder ved siden af og ned i flammer eller gløder, bliver det ødelagt. Dette idiom vil blive medtaget i Jysk Ordbog under *skydsel*, måske med henvis-

ning fra *skride*. Her må også nævnes (6) *være på de høje nagler*, der kan betyde 'have anspændt økonomi'; en anden betydning ses i afsnit 3.6.

3.2. KLOGSKAB vs. DUMHED; NARAGTIGHED; AKAVETHED

Vedr. KLOGSKAB findes adskillige idiommer med *ræv* som basisleksem, hvoraf nogle også kendes i rigsdansk, fx (7) *han er en snu ræv*, om en snedig person, (8) *han har en ræv bag øret* = 'er ikke alvorlig, men spørger el. narrer'; kun dialektal er derimod (9) *han er foret med rævebælg*, ligeledes om en listig person³. *Ræv* kan her ligesom *rose* ovenfor ses som et leksem med symbolsk funktion, et kulturelt symbol. I (10) *de tobenede ræve er de værste*, ses et eksempel på *polysemi*, idet *ræv* her fast betegner en person, og leksemet *ræv* tildeles i både ODS og DDO en betydning 2 'snu person'. Dette vælger Jysk Ordbog formentlig også, endskønt Feilberg blot nævner den figurative brug under ordets almindelige betydning.

Der kan her nævnes flere simile / analogier, der understreger, at afstanden mellem metafor og analogi kan være temmelig lille, fx (11) *han er snu som en ræv*, (12) *han er så lodden som æ ræv* (= ræven, vestjysk), om en upålidelig person. Brugen af *lodden* i betydningen 'snu' kan udstrækkes til andre dyr, fx (13) (han er) *en lodden hund*, idet også leksemet *hund* viser polysemi, jf. *hund 2* i Jysk Ordbog.

Mht. begrebet DUMHED har både dialekter og rigsdansk et betragteligt antal figurative enheder; jyske dialekter har fx idiom (14), *være på fårenes side*, der er medtaget i Jysk Ordbog (artiklen *får 1*, efter planke). Desuden (15), *ikke være ret hugget af træet*, som ikke er medtaget under verbet *hugge*, men sandsynligvis vil blive nævnt i artiklen *træ* til sin tid. Et tilsvarende billede af ufærdig tilblivelse ses i ordforbindelsen (16), *han er kun halvbagt*, hvor der i Jysk Ordbog under *halvbagt*, 'overført betydning' direkte oversættes til 'tåbelig. småtosset'. Et andet idiom er (17), *han kom af ovnen med kagen*, eller (17a) *han er kommet ud med kagen*; her betegner *kage* ikke kage som i dag, men derimod en slags hvidt brød, hvedebrød, som ved bagning i de gamle bageovne på gårdene blev taget ud længe før grovbrødene. Dette er medtaget i artiklen *kage 1.2* i Jysk Ordbog.

³ Jeg har tidligere behandlet brugen af *ræv* og andre "dyreleksemer" i idiommer (Arboe 2013b); enkelte nævnes i det følgende, fx subst. *får* i (14), *hare* (30).

At være helt klar over verdenshjørnernes placering var en nødvendighed tidligere; derfor kunne en persons dumhed omtales som (18) *hverken at kende sønder eller nør*, altså syd og nord. En selvironisk kommentar kunne være, at man (19) *havde fået sine tørv* (eller *heste*) *godt solgt*, dvs. var blevet narret; talemåden ikke medtaget i artiklen *hest* i Jysk Ordbog, men kommer formentlig med ved substantivet *tørv* eller verbet *sælge*. Udtrykket (20) *at køre med* (eller *sætte*) *de høje hjul forrest*, går tilbage til vogne i 1800-tallet og tidligere, hvor de mindste hjul skulle sidde forrest, som fx på diligencer og kareter, for at vognen kunne dreje uden at vælte. At sætte de høje hjul forrest ville være særdeles uhensigtsmæssigt; men hvis man gjorde det på trods af al fornuft, kunne det ses som tegn på overlegenhed, jf. Jysk Ordbog i artiklen ¹*hjul 1*. Engelsk har et lignende idiom, (20a) *to put the cart before the horse*, dvs. sætte vognen foran hesten, om at gøre tingene på en akavet eller bagvendt måde; parallelle udtryk hertil findes i flere sprog (Piirainen 2012:423f.).

3.3. LANGSOMHED, DOVENSKAB vs. HURTIGHED, HÅRDT ARBEJDE

Mht. LANGSOMHED og DOVENSKAB kan først nævnes den figurative enhed (21) *han vil ikke gerne stivne skagle*, med betydningen: vil ikke tage ordentlig fat, vil ikke arbejde hårdt. Skal dette ikke være semantisk helt opakt⁴ eller uigennemskueligt i dag, må man have den leksikalske oplysning med, at billedet er taget fra brug af hesteforspand tidligere: hesten skulle stramme skaglen, dvs. træk-rebet til hamlen, ellers trak den ikke vognen, ploven eller andet redskab fremad, men lod den anden / de andre hest(e) i forspandet tage slæbet. Det samme billede ligger bag det tilsvarende idiom (22) *vil helst gå med linde skagler*, egentlig: have slappe trækreb til hamlen. Idiomet (23) *han rider ikke den dag, han sadler*, er også kendt fra rigsdansk og derfor ikke medtaget i Jysk Ordbog. Derimod er (24), *springe i to kilder på en gang*, dvs. grave to brønde⁵ samtidig, nævnt under ¹*kilde 2*; brøndgravning var ofte et farligt arbejde, så man måtte koncentrere sig om én brønd ad gangen og evt. arbejde langsomt for at undgå nedskridning af

⁴ Jf. Cruse (1986:39) om 'degrees of opacity', grader af uigennemskuelighed.

⁵ Det gamle, dialektale ord *kilde* = 'brønd' er nu helt ertstattet af *brønd*; talemåden findes ikke med ordet *brønd*.

brøndens sider. I (25), *ikke løbe i syv kålgårde på en gang*, udtrykkes langsomhed ved at fremhæve, at en person i hvert fald ikke gør flere ting samtidigt.

HURTIGHED og HÅRDT ARBEJDE kan udtrykkes ved flere idiommer, der har hjul og hjuldele eller hest og vogn som basisleksemer, fx (26) *køre på egerne* og (27) *så hurtigt som kæp (eller kæpper) i hjul*. Det sidstnævnte er også kendt fra rigsdansk, og begge er nævnt i Jysk Ordbog, artiklerne ¹*ege* hhv. ¹*hjul 1* (dog kun som henvisning til behandling under subst. *kæp*). Desuden kan nævnes (28), *rende (eller) løbe som et vindt (eller skævt) hjul* = 'fare af sted', ligeledes i artiklen ¹*hjul 1*. I (29) *han måtte stinde skagler* = 'han blev nødt til at tage fat, arbejde hårdt', er udgangspunktet det samme som i (22). I (30) *er du sådan en hare* er basis derimod et dyreleksem, der medfører polysemi: Jysk Ordbog har ved subst. *hare* oprettet en betydning 2 = 'kryster'. Denne indgår også i en af betydningerne af talemåden (31a) *have et hareskind for måsen*. Videre er også idiommet (31) *stikke en kæp i hjulet* ('standse, bremse (nogen/noget)') velkendt i dialektale varianter, jf. (31a) *sæt en kjæp i hywle* (nordjysk), (31b) *stek en kjæp i æ hywl* (vestjysk) – foruden i andre sprog, jf. (31c) *put a spoke in the wheel* (engelsk), (32d) *den Stock ins Rad stecken* (tysk, nu forældet)⁶.

3.4. ØDSELHED vs. NÆRIGHED, GERRIGHED

Som eksempel på ØDSELHED kan nævnes den figurative enhed (32) *slå alt sit smør i bagetruget*, dvs. bruge det til baging, brød; dette kan ses som ødselt, hvis smørret kunne bruges bedre på anden måde, fx ved salg. I (33-34) er det igen dyreleksemer, der er basis for talemåden (33), *dit får har nok fået to lam* = 'du regner nok med at få meget uld senere, siden du ødsler sådan med den nu'; en variant er anført i artiklen *får 1* i Jysk Ordbog. Udtrykket kunne bruges (ironisk) over for en person, der ikke gik barfodet, men i strømper og derved sled disse hurtigt op, hvilket blev betragtet som ødselt i datidens selvforsyningsamfund. Baggrunden er den samme for (34), *du stoler nok på, at ronnevædderen er levende endnu* (nemlig så der kan komme lam nok til, at

⁶ Også afbildet som et blandt mange ordsprog på Pieter Brueghels maleri *De hollandske ordsprog* tilbage i 1559 (kan betragtes via Google).

du kan få uld til mange strømper), eller med andre ord: du er en optimistisk ødeland. Derimod er (35), *hun giver ikke sin sødmælksost bort* et udtryk for NÆRIGHED eller GERRIGHED, idet denne ostetype var bedre og sjældnere end andre. Tilsvarende gælder (36), *holde (eller spare) på skillingen*. Dette idiom har i midtvestjysk varianten (37), *håll· å æ skjelling å la æ då·ler rul· (eller spreng·)*, dvs. holde på skillingen og lade daleren rulle (springe), jf. *daler 1* i Jysk Ordbog; en talemåde også kendt fra rigsmål om folk, der er dårlige økonomer.

Hermed er de 4 semantiske felter med antonyme begrebspar behandlet; vi vender os nu mod de 4 semantiske felter med komplementære begrebspar.

3.5. HELD vs. UHELD

Vedr. HELD har jeg ikke fundet talemåder ud over det i afsnit 1 nævnte idiom, (4) *hans plov går rigtigt* = 'han tjener mange penge', der kan ses som udslag af held, skønt vel snarere dygtighed. Vedr. UHELD kan først nævnes det vestjyske (38) *å æ å·sk å i æ il'* = af asken og i ilden, med betydningen: 'fra noget ubehageligt til noget værre'. Tilsvarende kendes fra rigsdansk, og talemåden er derfor ikke anført i Jysk Ordbog, hverken under ¹*aske* eller *ild 1.5*, der ellers har et stort udvalg af de mange talemåder med ordet. Dernæst må nævnes (39) *af dynen og i halmen*, der vedrører de gammeldags sengesteder med sengehalm i stedet for madras; betydning: 'fra noget godt til noget dårligt', i Jysk Ordbog nævnt i begge artiklerne ¹*dyne 1* og ¹*halm*. I (40), *det er for sent at kaste kilden (eller brønden) til, når barnet er druknet*, udtrykkes en kritik af manglende opmærksomhed eller forudseenhed, naturligvis dermed også en opfordring til generelt at være mere forudseende for at undgå ulykker. I Jysk Ordbog anføres under ¹*kilde 2* en variant fra Vendsyssel: *de æ for sint å ful tsjå'len nær bå'nt æ dru'nt* = 'det er for sent at fylde kilden (brønden), når barnet er druknet'.

3.6. MUNTERHED, GODT HUMØR vs. VRANTENHED, VREDE

Vedr. MUNTERHED findes her (som i afsnit 3.3) figurative enheder med substantivet *skagle* som basis, jf. (41) *slå til skaglerne* = 'more sig, have det fornøjeligt'. Dette har udgangspunkt i, at hesten kunne finde på at sparke til en

af skaglerne, når den ikke ville trække mere eller ikke trække så hårdt længere; tilsvarende gælder (41a) *træde over skaglen* = 'begå et fejltrin'. Metaforen fik fornyet liv, da den i en lidt (og fejlagtigt) ændret udgave, (42) *slå en skagle*, blev brugt af John Mogensen omkring 1970 i sangen *Fut i fejlemøget*, her med betydningen 'blive beruset', jf. linjen: (42a) *når han har været på kroen at slå en skagle*, dvs. har drukket, så han er 'syg' dagen efter.

Netop beruselse er i øvrigt et område, hvor der er dannet mange metaforer eller konventionelle figurative enheder, fx (43) *han har fået målt skæppen fuld*, (44) *han har fået piben tændt*, (45) *han har vadet sin støvle over*, (46) *få blus på*, (46a) *have blus på lampen*, alle med betydningen '(meget, temmelig) beruset'. De er nærmere kommenteret i Arboe (2013a), og (43-45) vil formentlig blive medtaget i ordbogsartiklerne for *skæppe*, *pibe*, *støvle*, ligesom (46) allerede er anført i artiklen *blus 1*. Også udtrykket (47) *være på de høje nagler* kan betyde 'være meget beruset'; i afsnit 3.1 blev det nævnt som anvendt om en anden type opskruet situation, nemlig anspændt økonomi, tæt på fallit. Begge betydninger vil formentlig blive anført under subst. *nagle* i Jysk Ordbog. For tilstanden dagen efter rusen har dialekterne idiomet (48) *have tømmermænd* ligesom i rigsmål, desuden (49) *have en smed i panden*. Det kan tilføjes, at den nedertyske dialekt i Westmünsterland (over mod den hollandske grænse) har en tydeliggørende parallel til førstnævnte idiom her, jf.: (48a) *he häff de Timmerlöö in'n Dackstohl* = 'han har "tømmerfolk" i tagetagen' (Pirainen 2000.2:390).

Inden for anden del af dette semantiske felt, VRANTENHED, VREDE, er der registreret færre figurative enheder; dog kan nævnes: (50) *han er kommet baglæns i trøjen i dag* = 'er gnaven, sur', svarende til det rigsdanske (51) *har fået det forkerte ben ud af sengen*. Desuden foreligger to idiommer, der begge har subst. 'peber' som basisleksem: (52) *de maler peber* = 'er vrede, hidsige', og (53) *en kjole fuld af peber*, brugt overført om en hidsig, opfarende kone. I sidstnævnte eksempel kan det (jf. afsnit 2, indledningen) diskuteres, om basisleksemet er *kjole* eller *peber*; idiomet er medtaget i Jysk Ordbog i artiklen ¹*kjole 1.1* og vil måske også blive nævnt i artiklen *peber* senere.

3.7. SELVISKHED vs. SAMARBEJDE

Mht. SELVISKHED findes de figurative enheder (54) *hygge sine egne kartofler* og (55) *mele sin egen kage* også i rigsdansk; de er dog så velkendte i

dialektale varianter, at de er medtaget i Jysk Ordbog, jf. ¹*hygge 1* hhv. *kage 1.2*. En tredje, (56) *rage sammen til sin egen potte* (dvs. gryde), ses derimod kun i jysk. Vedr. SAMARBEJDE kan nævnes idiomet (57), *trække på samme* (eller *lige*) *hammel*, der har samme udgangssituation som idiomerne med *stivne skagle*, *stinde skagle* (22, 29) i afsnit 3.3: begge (eller alle) heste i et forspand skal trække lige meget i skaglerne / hamlerne, for at resultatet kan blive bedst muligt. Fælles for disse talemåder er, at de står i fare for at blive forældede og glide ud af sproget, efterhånden som de pågældende redskaber mv. ikke har været i brug i årtier; *trække på samme hammel* høres/ses dog stadig i officielt sprog, fx i Danmarks Radio.

3.8. LIV vs. DØD

Her angår de første eksempler nyt liv, graviditet: (58) *nu strammede kjolen igen om maven* (= 'nu var kvinden gravid igen'), er måske ikke helt etableret som konventionel figurativ enhed, men i hvert fald medtaget i Jysk Ordbog som citat i artiklen ¹*kjole 1.1*. En tidligere talemåde var (59), *hun går med en grøn ost i maven*, mens (60), *hun har brød i sin ovn*, måske kan blive det i Vendsyssel ud fra Ib Grønbechs brug af den i sangen *Kys mig, Connie, Connie, kys mig* fra ca. 1990; som forlæg har denne sang hit-melodien *Kiss me, honey, honey, kiss me* fra ca. 1960. *Brød i ovnen* i denne betydning er ifølge ODS-S rigsdansk slang fra 1950'erne; talemåden er ikke medtaget i Jysk Ordbog i artiklen *brød*. Talemåden (61) *hun har slået sin potte i stykker* er en let gennemskuelig metafor for, at kvinden har barslet.

I anden del af dette semantiske felt, DØD, finder vi i (62) den figurative enhed *at skride af skydselen*, som vi afsnit 3.1 mødte i betydningen 'gå fallit', jf. talemåden (6), men nu i betydningen betyde 'at dø'. (63) *han har fået en tørv* (eller *grøn tørv*) *over hovedet*, har været en udbredt talemåde for, at en person er død og begravet for lang tid siden. I (64), *at sætte eller stille træskoene*, er basisleksemet *træsko* ligesom i afsnit 1 (jf. (4), *gå baglæns af træskoene*), men her i et idiom med mere definitiv betydning. Denne talemåde bruges også i rigsdansk, men kan måske føles lidt gammeldags. I hvert fald har den daværende statsminister, Helle Thorning-Schmidt, for et par år siden peppet den op til en mere moderne version, da hun til en kritik af regeringens mange reformforslag bemærkede, at man ikke havde tænkt sig at standse hermed, nej, (64a) *det bliver ved til den dag, blondinen stiller stilet-*

terne! Et godt eksempel på ny udfyldning af et velafprøvet idiom; viser, at et gammelt talemådeforlæg kan moderniseres og bruges i nutidig offentlig kommunikation.

4. Afslutning

Undersøgelsen har vist, at man ud fra den foretagne opstilling af otte semantiske felter, organiseret som modsætningspar, kan give et godt indblik i, hvilke leksemer der ofte danner basis for konventionelle figurative enheder i form af især idiommer som metaforer og simile/analogier, i nogle tilfælde med polysemi og ny underbetydning af leksemet til følge. Som nævnt i afsnit 2.2 er idiommerne fundet inden for otte sagområder (gården og dens redskaber, køkken etc.), og inddrages disse eksplicit, kan basisleksemerne for de undersøgte konventionelle figurative enheder vises i en oversigt som i Figur 2.

Figur 2: Oversigt over basisleksemer i de konventionelle figurative enheder i afsnit 3.1-3.8: SEMANTISKE FELTER 1)–8) (antonyme begrebspar 1-4, komplementære 5-8), SAGOMRÅDER a) – h).

SAGOMRÅDE → SEMANT. FELT ↓	a) gård: redskaber, tilbehør mv.	b) vær- sted	c) dyr mv.	d) køkken- redskaber mv.	e) in- ventar mv.	f) be- klæd- ning	g) penge	h) geo- grafi
1) RIGDOM vs. FATTIGDOM	plov	nagler		skydsel	kiste	træsko	skilling	
2) KLOGSKAB vs. DUMHED	tørv vogn	(økse) hjul	ræv hund hest får	ovn				syd nord
3) LANGSOMHED vs. HURTIGHED MV.	skagle (sadel) kilder kålgårde skagle	eger kæp hjul	(hest) hare- skind					
4) ØDSELHED vs. NÆRIGHED			får, lam vædder	bagetrug sødmælks- ost			daler skilling	

5) HELD VS. UHELD	(plov) kilde			aske ild	dyne (senge-) halm			
6) MUNTERHED VS. VRANTENHED	skagle skæppe	nagler		blus peber	lampe seng	pibe støvle trøje kjole		
7) SELVISKHED VS. SAMARBEJDE	kartofler hammel			kage	potte			
8) LIV VS. DØD	tørv			brød ovn skydsel	potte	kjole træsko		

Skemaet omfatter altså 64 felter, der kan betegnes fra 1a) til 8h). Heri markerer fx placeringen af subst. *plov* øverst i felt 1a), at der med *plov* foreligger et idiom vedr. RIGDOM, nemlig *hans plov går rigtigt* = 'han tjener mange penge', jf. idiom (4) afsnit 3.1. Placeringen af subst. *skydsel* nederst i felt 1d) refererer parallelt hermed til et idiom vedr. FATTIGDOM, nemlig det vestjyske (5) *han skræj å æ skjøssel* = 'han gik fallit' i samme afsnit. Og så fremdeles for de fleste af leksemerne i skemaet. Parentes omkring et leksem, fx (*økse*) nederst i felt 2b), markerer dog, at ordet ikke direkte indgår i et idiom, men er underforstået i det: i afsnit 3.2 vedr. DUMHED implicerer talemåden (16) *han er ikke ret hugget af træet* brug af økse eller lignende.

Ud fra skemaet kan man let finde frem til de idiomer mv., der er omtalt i de foregående afsnit; fx ser man hurtigt, at *støvle* i felt 6f) refererer til afsnit 3.6 vedr. MUNTERHED, jf. talemåden (45) *han har vadet sin støvle over* ('han er beruset').

Princippet i inddelingen ovenfor kan videreføres ved kortlægning af andre konventionelle figurative enheder fra flere semantiske felter og sagsområder; dette vil kunne bidrage til en systematisering af dele af idiominventaret, både det dialektale og det rigsdanske. Flere af de nævnte metaforer og simile er på vej til at glide ud af sproget, efterhånden som de ting og redskaber, de refererer til, ikke har været i brug i mange årtier.

Litteratur

Ordbøger

- DDO = *Den danske ordbog*. 1-6 (2003-2005). København: Gyldendal.
- Feilberg = Feilberg, H.F. (1886-1914), *Ordbog over jyske almuesmål*. 1-4. København: Thieles Bogtrykkeri.
- Jysk Ordbog*: www.jyskordbog.dk (A-I), manuskript (J-KL), ordbogskartotek (KN-Å/AA). Peter Skautrup Centret for Jysk Dialektforskning, Aarhus Universitet.
- ODS = *Ordbog over det danske Sprog*. 1-28 (1918-1956). København: Gyldendal.
- ODS-S = *Ordbog over det danske Sprog. Supplement*. 1-5 (1992-2005). København: Gyldendal.

Anden litteratur

- Arboe, Torben (2008): Fraseologiske enheders indplacering i Jysk Ordbog. I: Ásta Svavarsdóttir et.al. (red.): *Nordiske Studier i Leksikografi*. 9. Reykjavik: Nordisk forening for leksikografi, 1-13.
- Arboe, Torben (2010): Faste forbindelser, med og uden tryktab. I: Harry Lönnroth och Kristina Nikula (red.): *Nordiska studier i lexikografi*. 10. Tammerfors: Nordiska föreningen för lexikografi, 86-102.
- Arboe, Torben (2013a): En- og flertydige idiom (metaforer) i jysk og rigsdansk. I: Simon Borchmann (et al.) (red.): *Gode ord er bedre end guld. Festskrift til Henrik Jørgensen*. Århus: Aarhus Universitet, 43-51.
- Arboe, Torben (2013b): Køb ikke katten i sækken – om idiom (især jyske) og deres behandling i ordbøger mv. I: Inger Schoonderbeek Hansen (et al.) (red.): *14. Møde om Udforskningen af Dansk Sprog*. Århus: Nordisk, Aarhus Universitet, 29-41.
- Cruse, D.A. (1986): *Lexical Semantics*. Cambridge: Cambridge University Press.
- Croft, William & D. Alan Cruse (2004): *Cognitive Linguistics*. Cambridge: Cambridge University Press.

- Dobrovol'skij, Dmitrij & Elisabeth Piirainen (2005): *Figurative Language. Cross-cultural and Cross-linguistic Perspectives*. Amsterdam (etc.): Elsevier.
- Lakoff, George (1987): *Women, Fire and Dangerous Things*. Chicago/London: The University of Chicago Press.
- Lakoff, George & Mark Johnson (1980/2003): *Metaphors We Live By*. Chicago/London: The University of Chicago Press.
- Piirainen, Elisabeth (2000): *Phraseologie der westmünsterländischen Mundart*. 1-2. Baltmansweiler: Schneider Verlag..
- Piirainen, Elisabeth (2012): *Widespread idioms in Europe and beyond*. New York (etc.): Peter Lang.
- Skautrup, Peter (1927-1979): *Et Hardsyssemål*. I-II. København: Universitets-Jubilæets Danske Samfund.
- Ømålsordbogen. En sproglig-saglig ordbog over dialekterne på Sjælland, Lolland-Falster, Fyn og omliggende øer. Tillægsbind* (1992). København: C.A. Reitzels Forlag.

Torben Arboe
ordbogsredaktør, vidensk. medarb., cand.mag.
Peter Skautrup Centret for Jysk Dialektforskning
Institut for Kommunikation og Kultur
Aarhus Universitet, bygn. 1483, 6.
Jens Chr. Skous Vej 4
DK-8000 Aarhus C
jysta@dac.hum.dk

Dansk og norsk i norske ordbøker

Sturla Berg-Olsen & Daniel Gusfre Ims

This article looks at the etymologies of words of Danish provenance in two one-volume dictionaries covering the two written standards of Norwegian. Our conclusion is that the etymologies often fail to indicate explicitly the Danish element, especially in Bokmål, where a number of lemmas have two alternative forms – one with roots in Old Norse and one from Danish. We recommend a gentle revision of the etymologies to amend this.

1. Innleiing

Utgangspunktet for denne artikkelen er ei undersøking av korleis etymologien til ord av dansk opphav og ord som er komne inn i norsk via dansk, er behandla i dei to ålmenndordbøkene *Nynorskordboka* og *Bokmålsordboka*. Det er eit velkjent faktum at dansk har hatt etter måten stor påverknad på norsk ordtilfang; for bokmålets del er dansk eit kjerneelement som framleis i stor grad ligg til grunn for målforma, trass i over hundre år med fornorsking. På nynorsksida var det lenge eit eksplisitt mål å minimere innslaget av danske ord, men likevel er den danske påverknaden openberr også i den målforma.

Undersøkinga viser at den etymologiske behandlinga av ord med tilknytning til dansk i dei nemnde ordbøkene er usystematisk og i nokre tilfelle reint villeiande. Rett nok er tilhøvet mellom norsk og dansk også i røynda ofte alt anna enn klart, og dialektdata kompliserer stundom biletet, men brukarane av ordbøkene fortener likevel ryddigare og meir konsekvente framstillingar av ordhistoria på dette feltet. Vi kjem med nokre konkrete framlegg til korleis ein kan oppnå dette med nokså enkle grep.

2. Materiale

2.1. Bokmålsordboka og Nynorskordboka

Bokmålsordboka (BOB) og *Nynorskordboka* (NOB) er eigde av Universitetet i Oslo og Språkrådet i fellesskap.¹ Dei kom ut på papir første gong i 1986, og sidan byrjinga av 1990-talet har dei vore fritt tilgjengelege på nettet. Dei to ordbøkene er autoritative kjelder til offisiell rettskriving og bøyning for bokmål og nynorsk, og for Språkrådet er dei viktige kanalar for formidling av dei gjeldande normene til ålmenta. Ordbøkene rettar seg primært mot den jamne norske morsmålsbrukaren, og dei er mykje bruka mellom anna i skolen og i den offentlege forvaltinga. I 2015 hadde nettutgåva av BOB 50 millionar søkjetreff og nettutgåva av NOB 38,8 millionar søkjetreff, ein auke på ca. 25 % frå året før (Språkrådet 2016).

Dei etymologiske opplysningane i BOB og NOB er jamt over nokså sparsame, noko som ikkje er urimeleg i lys av den breie målgruppa og av at det var klare rammer for plassen da papirutgåvene blei utarbeidde. Vi meiner likevel at det er viktig å setje av plass til etymologi også i ålmenndordbøker som desse. Interesse for ordhistorie er høg i Noreg, til dømes kan det nemnast at de Capronas *Norsk etymologisk ordbok* (2013) har vore inne på bestseljarlistene. Samstundes er det eit poeng at dei etymologiske opplysningane ikkje berre er sannferdige, men også lettfattelege.

2.2. Nokre tal

Ifølgje teksten på omslaga til dei nyaste papirutgåvene inneheld NOB ca. 90 000 lemma, medan BOB inneheld ca. 65 000 lemma. Søk i databasane som ligg til grunn for nettversjonane av ordbøkene, viser at forkortinga *da.* (dansk) er nemnd i etymologifeltet i 341 artiklar i NOB og 209 artiklar i BOB. Dette i seg sjølv seier kanskje ikkje så mykje, men det ser påfallande lite ut når ein samanliknar med kor mange artiklar det er som har etymologifelt med referansar til lågtysk (over 2000 i begge bøkene), tysk

¹ Universitetet i Oslo har vedtatt å avslutte verksemda si når det gjeld språksamlingar, inkludert engasjementet i BOB og NOB. Dette ansvaret vil bli overført til Universitetet i Bergen i 2016.

(ca. 1900 i NOB, over 2000 i BOB) og engelsk (ca. 1800 i NOB, ca. 1400 i BOB). For å kunne seie sikkert om det er slik at fleire lemma burde ha referanse til dansk, skal vi sjå nærare på ein del enkeltord som representerer ulike typar ord i norsk som på den eine eller andre måten har tilknytning til dansk.

2.3. Lemma av dansk opphav

Om vi tar for oss lemma som utan tvil er komne inn i norsk frå dansk, og granskar etymologifeltet for desse lemmaa i BOB, ser vi at det danske opphavet av og til er kommentert, medan det andre gonger ikkje er nemnt i det heile. Eit døme på det første – at dansk står nemnt – er *edder*:

edder *-en* el. *-et* (da. form av *eiter*) [...]²

Her kan ein merkje seg at det ikkje står at *edder* er eit lån frå dansk i norsk – berre at *edder* er ei dansk form av lemmaet *eiter*. Det er ikkje heilt klart korleis ein skal tolke dette, men det er nærliggjande å tenkje at ei dansk form som *edder* i utgangspunktet ikkje blir sett på som framand. Jamfør denne formuleringa om opphavet til bokmålet i Språkrådets retningslinjer for normering (Språkrådet 2015:9): ”Skriftspråket bokmål byggjer på fornorskinga av det danske skriftspråket som vi hadde i landet fram til slutten av 1800-talet [...]”. Dansk ligg altså til grunn for bokmålet, og frå denne synsvinkelen er ikkje ei form som *edder* innlånt. På den andre sida har fornorskinga ført til at ein òg har forma *eiter*, men desse to formene er såpass ulike at dei neppe bør behandlast som eitt lemma. Eit anna døme er artikkelen om (*h*)*verken* i BOB:

II verken el. **hverken** [...] konj (fra da. *hverken*, gda *hwærki(n)*, jf norr *hvár(t)ki* ’ikke noe av to’) [...]

Også her er det danske opphavet ukontroversielt, men vi merkjer oss at formuleringa er noko annleis her enn for *edder*: ”fra da.” indikerer strengt tatt at dette er eit lån frå dansk i norsk. Den nynorske forma *korkje* er derimot ei av

² Her og nedanfor siterer vi av praktiske omsyn frå artiklane slik dei ligg føre i nyaste papirutgåve av BOB og NOB.

fleire lydrette norske former av norrønt *hvár(t)ki*. Slår vi opp på *korkje* i den store og stort sett deskriptive *Norsk Ordbok* (NO), finn vi interessant nok også forma *værken*, her oppgjeven som målføreform med heimfestinga ”vanleg, særleg i bymål”. Det er uklart kva grunnen til denne føringsmåten er; kanskje ein under redigeringa av bind 6 av NO ikkje meinte at *verken* burde få status som eige lemma, men likevel hadde behov for å vise at denne forma fanst i målføra? I bind 12 finn vi uansett ein eigen artikkel på *verken*, så det er mogleg konklusjonen blei at føringsmåten i bind 6 var mindre heldig.

Vidare skal vi sjå på eit par døme der ord eller former av dansk opphav står oppførte i BOB utan eksplisitt referanse til dansk. Det første dømet er *heder*:

heder -en (norr *heiðr*) [...]

Det verkar nokså opplagt at bokmål *heder* er den same forma som dansk *hæder* (ifølgje ODS òg skrive med *e* i eldre dansk). Men det er i beste fall problematisk å berre gje den norrøne forma *heiðr* i etymologien utan vidare forklaring på tilhøvet mellom *heiðr* og *heder*. Ein kunne til dømes føre også gammaldansk *hedher* her, jamfør føringa av ei gammaldansk form i artikkelen om (*h*)*verken*. Eit anna ord der dansk burde vore nemnd, er *får*. BOB skriv:

II får n1 (norr *fær*) [...]

NOB viser til dansk *får* i forleddsartikkelen **fåre-**, og NO gir eit vidare etymologisk utsyn i sin artikkel **III får**: ”ett da. *får* ’sau’; jfr gno (austl) *får*, gno (vestl) *fær*”. Grunnordet *får* finn ein i norsk berre i litterære – til dømes bibelske – kontekstar. Som umarkert målføreord finst det ikkje i det heile, *korkje* i forma *får* eller *fær*. Dette er da eit klart lån frå dansk (eventuelt òg med påverknad frå svensk, sjølv om det er mindre truleg).

2.4. Lemma med parallelle former i *Bokmålsordboka*

Det er ei kjend sak at bokmål i mange tilfelle har to former av same ordet – ei med tilknytning til dansk og ei med bakgrunn i norske målføre – ofte med ulike stilistiske valørar og av og til også noko ulike tydingar. Eit spørsmål

for seg er i kva grad slike par skal førast som variantar av same lemma eller lemmatiserast kvar for seg; her finst det ulik praksis i ordbøkene, jamfør òg diskusjonen i Lauvstad (2012). I dei tilfella der to parallelle former blir førte saman, melder spørsmålet seg om kva ein gjer med etymologien. Vi skal sjå på nokre døme med utgangspunkt i BOB.

II lav el. **III låg** a1 (da. *lav*, norr *lágr*, besl med *ligge*) [...]

I dette tilfellet ser vi at den danske forma *lav*, som ligg til grunn for bokmål *lav*, er nemnt, ved sida av norrønt *lágr*, som ligg til grunn for *låg*.³ På sett og vis er det altså gjort greie for begge variantformene, men det ville vere nyttig for interesserte brukarar å få meir eksplisitt informasjon om tilhøvet mellom dei. Når etymologiane i BOB er heller stutte, er det motivert av plassomsyn i papirutgåva. Ved ein revisjon av BOB bør ein vurdere å vie noko meir plass til etymologiane, også fordi ein i dag ikkje lenger er bunden til papirformatet.

melk el. **mjøl**k -a el. -en (norr *mjolk*, da. *melk*) [...]

Her ser vi noko av det same – den danske og den norrøne forma er lista opp, men utan vidare kommentar.⁴ I tillegg kan ein stusse på at *melk* står som dansk form i staden for *mælk*, men ODS viser at ordet i eldre dansk òg blei stava med *e*. Det kan nemnast at NO behandlar *melk* og *mælk* som målføreformer av lemmaet *mjøl*k, utan at dansk er nemnd i etymologien der.

I andre tilfelle har BOB lemma med to parallellformer utan at dansk er nemnd:

fram el. **frem** adv (norr *fram* 'god, gjev') [...]

I lek el. **II leik** m1 (norr *leikr*) [...]

³ Merk at ein ikkje kan seie at den norrøne forma *lágr* ligg til grunn for dansk *lav*, ettersom norrønt per definisjon er vestnordisk.

⁴ Eit forvirrande moment her er òg at rekkjefølgja på formene i etymologifeltet ikkje speglar rekkjefølgja til oppslagsformene, men dette er ikkje noko som går att elles i ordboka.

Her er det berre den norrøne forma som er oppgjeven i etymologien. Ein kan argumentere for at former nær opptil norrønt *fram* og *leikr* i siste instans også ligg til grunn for dei danske formene *frem* og *leg*, men å berre oppgje norrønt her utan vidare kommentar blir likevel vel knapt.

Paret *lek/leik* involverer nokre kompliserande faktorar; *lek* er nemleg det ein kan kalle ei hybridform. Monoftongen *e* samsvarar med eldre dansk uttale og med leseuttale av dansk slik han blei praktisert i Noreg. Den historisk stemde utlyden *g* i dansk er skifta ut med den ustemde *k*-en i den nedervde norske forma (Myrvoll 2006:181–182). I norske målføre med monoftongering (mellom anna Sør-Østfold, Østerdalen og fleire område i Sør-Trøndelag) er *lek* den lydrette refleksen av norrønt *leikr*, men dette faktumet rokkar ikkje ved at det er dansk som ligg bak forma *lek* i standardspråket bokmål.

2.5. Lemma der ”dansk” står i etymologien berre i *Nynorskordboka*

Sidan bokmål har eit tettare samband med dansk enn nynorsk, kan det verke noko paradoksalt at det er i NOB vi oftast finn ”dansk” i etymologien (341 gonger mot 209 i BOB). Det danske opphavet er referert til på ulike måtar, som ”av da.”, ”frå da.”, ”gjennom da.” og ”truleg frå da.”. Dessutan finn vi refereransar til eldre steg av dansk, som ”g.da.” (gammaldansk), ”av eldre da.” og ”av forelda da.”. To døme på ord som har referanse til dansk opphav berre i NOB, er *allehelgensdag* og *bregne*.

allehelgensdag (da.) helgemess

bregne m1 (da.) plante av gruppe av familiar (i klassa *Filicinae*) som høyrer til karsporeplantene; burkne (1)

I BOB står dei slik:

allehelgensdag minnefest for martyrer og helgener; i Norge: første søndag i november; i den romersk-katolske kirke: 1. november

bregne f1 el. m1 (besl med dial *brake* ’einer’, sv *bräken*) karsporeplante av familier i klassen *Filicinae*

Vi ser at NOB i begge desse tilfella viser til nynorske synonym – *helgemess* og *burkne*. Kanskje kan dette vere noko av forklaringa på at ”dansk” oftare er nemnt i NOB enn i BOB; nynorsk har eit synonymt ord som er vanleg eller vanlegare, og den etymologiske tilvisinga får fram meir av bakgrunnen for dei ulike lemmaa.

Øg i BOB finn vi både *burkne* og *helgemess*, og sjølv om desse orda er mindre vanlege på bokmål enn på nynorsk, meiner vi det vil vere naturleg at BOB og NOB har med dei same etymologiske opplysningane. Gjennomgåande bør ”dansk” stå i etymologien i BOB alle stader der det står i NOB.

2.6. Lemma i *Nynorskordboka* der ”bokmål” står i etymologien

Det er interessant at vi i tillegg til referansane til *dansk* i NOB finn ein del døme på at *bokmål* står i etymologien. NOB har med ”bm” 181 gonger; mot-svarande viser BOB til *nynorsk* (”nyno.”) berre 41 gonger. Vi skal sjå nærare på nokre av desse orda som ifølgje NOB skal vere frå bokmål.

anbod n1 (av *I an-*, gj bm frå ty., lty) [...]

hurtig a2 (gj bm, da. frå ty.) [...]

Desse oppføringane reiser fleire spørsmål. For det første ser vi at både *anbod* og *hurtig* skal ha opphav i (låg)tysk, medan berre *hurtig* òg har referanse til dansk. Manglar det ikkje da ein referanse til dansk òg for *anbod*? Den direkte påverknaden frå lågtysk var over ei god stund før *bokmål* blei fastsett som nemning i 1929, og før dette skriftspråket for godt skilde lag med den danske rettskrivinga i 1907.

Kva vil det eigentleg seie at eit ord i nynorsk kjem frå bokmål? Er referansen til bokmål med i NOB fordi ord som *anbod* og *hurtig* har stått lenger i bokmålsordbøker enn nynorskordbøker? Det er openbert rett, for i nynorsktradisjonen var det lenge vanleg å unngå ein del ord med opphav i dansk og tysk. Likevel meiner vi ein må spørje seg om ikkje ein del av desse orda har vore vanlege i talemåla som ligg til grunn for nynorsk, før bokmål var etablert som ein eigen standard (det første tiåret under namnet *riksmål*).

Om vi granskar den skriftlege bruken av *anbod* og *hurtig* i nynorsk, ser vi at begge desse orda faktisk er bruka etter måten tidleg. Til dømes finn vi ein

annonse med denne ordlyden i avisa *Den 17de Mai* 1. juli 1911: ”Bed um anbod!”. Likeins finn vi ei setning som dette i avisa *Norske Intelligenssedler* 28. september 1916: ”So hev det vore tale um ein [sic] hurtig ruta til Bergen naar Raumabana kjem istand.”

Vi meiner med andre ord det er grunn til å hevde at både *anbod* og *hurtig* har vore etablerte i talemål som ligg til grunn for nynorsk, før det var tale om eit eige norsk bokmål som skilde seg frå dansk. Følgjeleg må lån som dette ha komme gjennom dansk til både bokmål og nynorsk, eventuelt direkte frå lågtysk til både dansk og norsk. Vi reknar det som sannsynleg at tilsvarande gjeld for fleire av orda som har ”bm” i etymologien i NOB. Som ein konsekvens av dette meiner vi fleire av etymologiane i NOB må vise til dansk i staden for eller i tillegg til bokmål. Det må likevel seiast at det kan vere vanskeleg å dokumentere påverknaden mellom bokmål og nynorsk på ein fullgod måte. Det kan tenkjast at ord som *anbod* og *hurtig* ikkje ville ha vore vanlege på nynorsk i dag om det ikkje var for at *anbud* og *hurtig* alltid har vore vanlege på bokmål. Tilsvarande kan ein hevde at ord som *trygd* og *søknad* nok ikkje hadde etablert seg i bokmål om det ikkje var for at nynorsk hadde desse orda.

3. Nokre prinsipielle vurderingar

Er det gale å vise til norrønt ved danske former? Myrvoll (2006) hevdar i ei melding av den etymologiske ordboka *Våre arveord* (Bjørvand & Lindeman 2000) at det er ”ein leid tendens i norsk målvitskap til å tøkeleggja saga og opphavet åt det norsk-danske målet i Noreg” (s. 174). Denne tradisjonen framstiller bokmål (*norsk-dansk* hos Myrvoll) ”som eit opphavlegt norsk mål som berre hev vorte utsett for påverknad frå dansk” (ibidem). Og vidare: ”Noko som sermerkjer *Våre arveord* i høve til mange andre framstillingar av norsk-dansken (t.d. *Bokmålsordboka*), er at det vert sagt beint fram at den og den ordformi ættar frå dansk” (s. 181).

Mot dette kan vi hevde at etymologien i BOB eigentleg berre seier kva den norrøne forma *er*; det står ”norr”, ikkje ”av norr” eller ”frå norr”. Brukarrettleiinga i BOB seier det slik: ”Når et hjemlig, usamensatt ord er belagt i norrønt, er den norrøne formen alltid oppført.” Spørsmålet blir da om vi

skal rekne dansk som "hjemlig", og ordforklaringa til *bokmål* i BOB talar for ei slik tolking:

bokmål [...] **2** fra 1929: betegnelse for den ene av Norges to offisielle målformer, utviklet fra dansk og dansk-norsk og påvirket av norsk talemål, t forskjj fra *nynorsk* [...]

Vidare kan vi hevde at dansk har utvikla seg frå former som i alle fall står norrønt nær, slik at norrøne former òg er interessante i framstillinga av opphavet til danske former (jamfør oppføringane i ODS).

Likevel meiner òg vi at oppføringsmåtane i BOB lett vil gi inntrykk av at former med klart dansk opphav har same utgangspunktet som andre former når det berre er vist til ei norrøn form. Dette blir sjølvsagt styrkt ved at det danske opphavet likevel er nemnt ved nokre etter måten få ord. På denne måten kan den jamne ordbokbrukaren lett tenkje at eit ord kjem frå norrønt, sjølv om det strengt tatt berre står kva form som fanst på norrønt.

Skulle vi lagt utgangspunktet hos Myrvoll til grunn ("norsk-dansken er ein dansk som hev vorte tilmåta norsk mål, og som ikkje minst hev teke upp mange lånord frå dette målet" (2006:174)), kunne ei ny løysing for etymologifeltet i BOB ha vore å gjere om lag det motsette av i dag – alltid opplyse om den danske forma og vise til den norrøne forma berre der vi veit at det er ei norsk form.

Vårt framlegg til løysing er mindre drastisk. I ordbøker som har med etymologiopplysningar, må òg dansk opphav komme fram på ein så gjennomført måte som mogleg. Likevel meiner vi tilvisinga til norrønt framleis er ordhistorisk relevant.

4. Framlegg til ei praktisk løysing

Det doble opphavet i dansk og norsk/norrønt i den norske ordhistoria bør komme tydelegare fram. Vi kan ikkje rekne med at dette er ålmennkunnskap.

I avsnitta nedanfor har vi framlegg til korleis nokre lemma kan ha opplysningar om både norrøne og danske former, medan andre framleis kan ha opplysningar berre om norrønt. Lemmaoppføringane i denne delen står følgjeleg ikkje i BOB og NOB i dag; det er berre våre framlegg. Vi legg vekt på

at framstillinga skal vere forståeleg for den jamne lesaren av ålmenndbøker som BOB og NOB, men prinsippa i framlegget bør kunne gjelde generelt for ordbøker med etymologiopplysningar.

4.1. Lemma med opplysning om både dansk og norrønt

4.1.1. Lemma med parallelle former

fram el. **frem** adv (*fram*: norr *fram(m)*; *frem*: da. *frem*)

Løysinga som er skissert her, svarar til den som i dag finst for *mjølkk/melk*, berre at rekkjefølgja her viser samanheng mellom dei moderne formene i bokmål og formene i høvesvis norrønt og dansk. I dette framlegget har vi dessutan poengtert kva former som har samband med høvesvis norrønt og dansk. Eit alternativ kunne vere å knyte referansane til norrønt og dansk direkte til høvesvis *fram* og *frem* med to ulike parentesar. Men BOB og NOB er først og fremst rettskrivingsordbøker, og vi meiner det vil vere uheldig med meir tekst mellom dei jamstilte rettskrivingsformene *fram* og *frem*. Sambandet med høvesvis norrønt og dansk kunne i dette tilfellet òg ha komme fram ved å ha same rekkjefølgja i etymologien som for dei noverande jamstilte formene. Løysinga med gjentakning i parenteser er vald for å kunne gjennomføre den same metoden òg i kombinasjonar med dobbelt opphav og hybridformer (som nedanfor).

4.1.2. Lemma med dobbelt opphav; hybridformer

hvete m1 (norr *hveiti*, da. *hvede*)

lek el. **leik** m1 (*lek*: norr *leikr*, da. *leg*; *leik*: norr *leikr*)

Hvete finst berre i denne hybridforma på bokmål – med monoftong som på dansk og ustemt konsonant som i søraustnorsk talemål. *Kveite* (eineform på nynorsk) var ei parallell form i bokmål fram til rettskrivingsreforma i 2005, medan *hveite* aldri har vore med i den offisielle rettskrivinga. Vi meiner

etymologien må ha med referanse til både dansk og norrønt, og vi har sett det eldste språksteget først.

Ei knapp og minimalistisk framstilling av etymologien til hybridformene er generelt ei utfordring, og i neste døme er det enda eit moment. *Lek/leik* er parallelle former, og i tillegg er *lek* hybridform. Som vi viste ovanfor i dømet med *fram/frem*, tek vi til ords for ei klargjerande framstilling i parentes. Ei knappare framstilling, til dømes ”(da. *lek*, norr *leikr*)” ville òg få fram at monoftongvarianten har eit tettare samband med dansk, medan diftongvarianten har eit tettare samband med norrønt. Ei slik framstilling kunne likevel òg lesast slik at *lek* berre har med dansk *leg* å gjere.

4.1.3. Lemma av dansk opphav

heder m3 (da. *hæder*, jf norr *heiðr*)

Jamvel om den danske forma i dette tilfellet har fortrent ei eventuell norsk form utvikla frå norrønt, meiner vi ei jamføring med norrønt vil vere relevant. For det første viser det kva forma var på eit eldre nærskyldt språksteg. For det andre kan det argumenterast for ein kontinuitet på ordnivå sjølv om det er brot på formnivå, slik Myrvoll (2006:175–176) drøftar:

[...] no. *heider* og da. *heder* [kann] segjast å vera same ord. Eit innlån av ordformi *heder* i norsk kann ein då skildra som at ordet HE(I)DER fær bytt ut formi si frå *heider* til *heder*. Fyresetnaden for ein slik analyse er at språkbrukarane koplår ordformene i dei tvo måli med einannan, noko som er mykje trulegt i eit tilfelle som no. *heider* = da. *heder*, og so av einkvan grunnen set seg fyre å nytta den danske formi av ordet i staden for den norske. Me fær då kontinuitet på ordnivå – ordet HE(I)DER vert verande i målet – men eit brot på formnivå; det hev ikkje lenger den nedervde formi.

For lemma av dansk opphav som ikkje finst i ei anna form i norrønt, vil det sjølvsagt berre vere relevant å vise til dansk åleine. Klarast blir dette med nyare lån, som *duppedit*.

4.2. Lemma med opplysning berre om norrønt

4.2.1. Lemma med berre norsk form

hei m1, fl (norr *heiðr*)

Denne oppføringa er som i dag. Ordet har opphav i norrønt og finst i dag berre i norsk form.

4.2.2. Lemma med ”fellesform”

hest m1 (norr *hestr*)

kaste v1 (norr *kasta*)

Lemma som *hest* og *kaste* er fellesformer på den måten at utvikling frå dansk eller beint frå norrønt ville ha gitt det same resultatet. Prinsipielt kunne slike lemma hatt med *dansk* i etymologien. Om vi hadde tatt med referanse til både dansk og norrønt, kunne vi vist at vi heller ikkje veit sikkert kva mål ordet er overført frå. Hadde vi derimot fullstendig lagt synet til Myrvoll til grunn (med dansk som utgangspunkt og lånord frå norsk), burde vi berre ha vist til dansk. Vår løysing seier altså indirekte at dei moderne formene i bokmål har opphav i norrønt med mindre noko anna er bevist.

4.3. Lemma i NOB og referansane til bokmål og dansk

hurtig (gjennom da. frå ty.)

I ein gjennomgang av etymologiane i NOB meiner vi det bør vurderast å byte ut *bokmål* med *dansk* i dei tilfella der ordet først kjem frå dansk. Påverknaden frå bokmål er utvilsamt viktig i utviklinga av ordtilfanget i nynorsk, men i tilfelle som *hurtig* og *anbod* meiner vi bokmålpåverknaden helst dreier seg om frekvensen, ikkje eigentleg om opphavet. Med frekvenspåverknad meiner vi at ein kan tenkje seg at *hurtig* og *anbod* hadde blitt heilt bytt ut med andre ord i nynorsk om det ikkje var for at dei heile vegen fanst i bokmål. På tilsvarande vis kan vi seie at mange lånord frå latin i dag blir

styrkte gjennom kontakt med engelsk. Men det er uråd å leggje dette omsynet til grunn så sant vi skal gjere arbeidet med etymologiopplysningane overkommeleg.

5. Konklusjon

Hovudpoenget vårt er at det trengst ei opprydding og systematisering av etymologifeltet når BOB og NOB skal reviderast. Særleg bør det dobbelte opphavet for bokmål i dansk og norrønt komme tydeleg fram i BOB. Etymologiane må vere baserte på det vi veit. Samstundes bør det ikkje vere for plasskrevjande oppføringar og ikkje for tidkrevjande revideringsarbeid; løysingsframlegga våre følgjer difor for det meste ei minimalistisk linje.

Litteratur

Ordbøker

BOB = *Bokmålsordboka* (2005). Boye Wangensteen (red.). 3. utgave. Oslo: Kunnskapsforlaget. <<http://www.nob-ordbok.uio.no/perl/ordbok.cgi?OPP=&bokmaal=+&ordbok=bokmaal>> (september 2015).

de Caprona, Yann (2013): *Norsk etymologisk ordbok. Tematisk ordnet*. Oslo: Kagge forlag.

NO = *Norsk Ordbok. Ordbok over det norske folkemålet og det nynorske skriftmålet* (1966–2016): Oslo: Det Norske Samlaget. <no2014.uio.no> (september 2015).

NOB = *Nynorskordboka* (2006). Marit Hovdenak et al. (red.). 4. utgåva. Oslo: Det Norske Samlaget. <<http://www.nob-ordbok.uio.no/perl/ordbok.cgi?OPP=&nynorsk=+&ordbok=nynorsk>> (september 2015).

ODS = *Ordbog over det danske Sprog*. <<http://ordnet.dk/ods>> (september 2015).

Annan litteratur

- Bjorvand, Harald & Fredrik Otto Lindeman (2000): *Våre arveord. Etymologisk ordbok*. Oslo: Novus.
- Lauvstad, Hanne (2012): Atypiske lemmaer i Det Norske Akademis Store Ordbok. Noen betraktninger om deres plass og funksjon i en moderne diakron og deskriptiv ordbok. I: Birgit Eaker, Lennart Larsson & Anki Mattisson (red.): *Nordiska studier i lexicografi 11. Rapport från Konferensen om lexicografi i Norden Lund 24–27 maj 2011*, 396–404.
- Myrvoll, Klaus Johan (2006): Kva er eit erveord i norsk-dansk? Noko um den nære ordsoga i *Våre arveord*. I: *Maal og Minne* 2:173–212.
- Språkrådet (2015): *Retningslinjer for normering av bokmål og nynorsk. Nynorskversjon*. <<http://www.sprakradet.no/globalassets/spraka-vare/norsk/retningslinjer-for-normering-nynorskversjon.pdf>> (november 2015).
- Språkrådet (2016): *Rekordmange søk i ordbøkene*. <<http://www.sprakradet.no/Vi-og-vart/hva-skjer/Aktuelt/2016/rekordmange-sok-i-ordbokene/>> (januar 2016).

Sturla Berg-Olsen
seniorrådgjevar, dr.art.
Språkrådet
Postboks 8107 Dep
NO-0032 Oslo
sturla.berg.olsen@sprakradet.no

Daniel Gusfre Ims
seksjonssjef, cand.philol.
Språkrådet
Postboks 8107 Dep
NO-0032 Oslo
daniel.ims@sprakradet.no

Synonymitis

Lars Brink

I try to show that there are far too many (alleged) equivalents per meaning variant in our bilingual dictionaries, in other words, (alleged) synonyms. I clarify the concepts of 'equivalent', 'false synonyms', 'genuine synonyms', and 'maximal synonyms', and state that, generally, our users only need one equivalent, the best, and that more than one is pedagogically and economically harmful. The L1 speaker only needs one expression for his concrete translation. He does not speak in synonyms ("my steed and horse"). The passive L2 speaker and the active L2 speaker only want to understand. One equivalent will suffice. They do not need to be informed of alternative expressions in their own mother tongue. Later, I sketch where it is useful to give several equivalents. And in the end I seek to answer why the one word ideal is not as popular as it ought to be.

1. Introduktion

For mange år siden skrev leksikografen Jørgen Bang i sin sprogklumme (Berlingske Tidende 22.6 1973) om den lumske lighed mellem svensk *stilig* og dansk *stilig*. Dansk har lånt ordet fra svensk, og vi opfatter det netop som et svensk lån, da det næsten kun bruges om svensk folkekarakter: *de stilige svenskere*, hvad der, på dansk, betyder 'de stive (dvs. formelle) svenskere'. Bang forklarer så, at den svenske betydning faktisk er:

1. "flot, smart, storslået, storstilet"

Denne kvadruplet er ikke vejet på guldvægt – og det skal være ham tilgivet, i en letløbende klumme – men når vi skriver ordbøger, *skal* vi veje på guldvægt. Og gør jeg dét, må jeg forkaste tre af dem:

Smart har noget moderigtigt i sig, som ikke ligger i sv. *stilig*, altså i selve ordets signaler. Konteksten kan evt. vise, at afsender finder det pågældende både 'flot' og 'smart', ligesom en kontekst kan vise, at en bestemt mand er smart, men dét indebærer ikke, at *mand* her betyder 'smart mand'. Be-

tydningen 'smart' signaleres ikke af *stilig* hhv. *mand*. *Smart* er særlig uheldigt, fordi det har en meget almindelig let negativ betydningsvariant: *en smart advokat* osv., mens sv. *stilig* er stærkt positivt.

Storslået er alt for "stort": Det kan fx slet ikke bruges om tøj, og om personer går det på den generøse karakter, mens sv. *stilig* går på den ydre fremtoning.

Storstilet er helt skævt: *Enorstilet plan* eller *etorstilet byggeri* siger intet om det æstetiske, men dét gør sv. *stilig*.

Den nye, glimrende Svensk-Dansk Ordbog (Kristensen m.fl. 2010) har:

2. *stilig* ~ *flot*; *elegant*

Og her godtager jeg gerne dubletten. Blot ville jeg foretrække komma eller skråstreg imellem de to, idet jeg mener, at svensk *stilig* kun har én betydning, som er et overbegreb for de to danske betydninger. Disse to danske betydninger er klart nok forskellige: *flot* betoner det imponerende og kræver et vist volumen, mens *elegant* betoner det æstetiske, uden krav om volumen, jf. at lille Allan Simonsen ikke var en *flot* fodboldspiller, men særdeles *elegant*. Det er velkendt, at ét ord på et sprog dækkes af et felt bestående af to ord på et andet sprog. Svensk har *kusin* dækkende 'fætter/kusine'. Sv. *kusin* betyder aldrig 'fætter', så sandt som sv. *hund* aldrig betyder 'hanhund', men altid dækker hele feltet. Hvis teksten viser, at det er en hanhund, er det med garanti ikke pga. morfemet *hund*. Vi vil jo heller aldrig sige, at *bil* kan betyde 'blå bil'.

2. Definitioner

Før vi går videre, må vi, naturligvis, igennem en terminologisk renselse:

En ÆKVIVALENT til et ord eller udtryk = 'et etableret udtryk i et andet sprog, der har samme denotation og konnotation'. (Hvis det fremmede udtryk fx er formet som en uetableret definition, er det ikke en ækvivalent).

DENOTATION er det samme som SANDHEDSBETYDNING og SANDHEDSKRITERIER, dvs. de betingelser, ordet stiller til virkeligheden for, at ytringen er sand. Med et forslidt eksempel: *Vovse*, *hund*, *køter* har samme sandhedsbetydning. De stiller ét eneste krav til virkeligheden for, at den

pågældende ytring er sand: At der er tale om dette bestemte dyr. Dyrets race, farve og popularitet har ingen indflydelse på ytringens sandhedsværdi.

Leksikografen bør i øvrigt gøre sig klart, at de fleste ord ligesom *hund* er appellativiske, dvs. kun stiller sandhedskrav til virkeligheden om, at *et eller andet medlem af en klasse* skal være involveret: *Svømme* denoterer en hvilken som helst svømmeakt (crawl, hundesvømning, med el. uden badehætte), *tyk* betegner en uangivet bredde mellem det velvoksne og det smældfede, *fortrinsvis* (*Koret synger fortrinsvis klassiske værker*) betegner en eller anden værdi mellem 'næsten udelukkende' og 'overvejende', *på* betegner en beliggenhed med kontakt på en flade, det være sig en flues på indersiden af en flaske eller indlandsisens tryk på grundfjeldet, *fordi* betegner, at der følger en eller anden årsag: morderens hånd, giftens virkning, et vindpust. Kun *proprier* og få andre, fx *nu* og *og*, denoterer ikke et uangivet medlem af en klasse. Med *Sokrates* betegner vi denne ene, afdøde mand. *Og* betegner, at noget sandt også gælder det følgende led. Her er der én relation, ikke en skov af åbne muligheder. De appellativiske ords åbne muligheder kan naturligvis indsnævres til én: *Hundén – mín crawl de sidste 10 m – af dén grund at osv.*

Rundt om sandhedsbetydningen sværmer så alskens KONNOTATIONER. De vigtigste er: 1. værdiladning. 2. stil (neutral, formel, uformel, højstemt, poetisk, arkaiserende, respektløs, fagsprogs-, Holberg-, dialekt- el. børnesprogs-præg osv.). 3. applikations-restriktion (*fjeld* er denotativt = 'bjerg', men begrænset til nordiske bjerge). *Dusin* = 'antallet 12', men kun om visse småtterier, aldrig om hjerteslag eller stjerner. *Dyr* appliceres især på pattedyr, sjældnere på fugle og fisk, næsten aldrig på bakterier). 4. Associations-sfæren: *Splitternøgen* giver kraftige associationer til samfundets nøgen-tabu, *adamskostume* mildner tabuet med let humor. Latin *pellere* og *pulsare* kan vist altid ombyttes i grundbetydningen, med samme denotation, men *pulsare* er mere intensivt og bruges især om gentagne slag, man "mærker" her tydeligere slagene. Tilsvarende *jage* og *jagte*. *Jehova* associerer uvægerlig til Jehovas Vidner, selvom afsender tænker i andre baner, ligesom *frihedskæmper* associerer til Danmarks besættelse, selvom det udmærket kan bruges, positivt, om alverdens partisaner.

Særlig stor associations-forskel er der mellem udtryk uden pålydende betydning og deres synonyme med tydelig pålydende betydning. (Pålydende betydning = bogstavelig betydning, videnskabeligt præciseret, se Brink

2010): *Film* er (på dansk) uden pålydende betydning, mens *levende billeder* dels fremmaner noget spændende dynamisk, dels filmens barndom. Tilsvarende *slåbrok* vs. *badekåbe* – *kodeks* vs. *lovbog* – *gammel* vs. *bedaget*.

Ved UÆGTE SYNONYMER forstår jeg – i al vaghed – udtryk, hvis denotationer ligner hinanden meget, men ikke falder sammen. Her har vi synonymordbøgernes buket af former:

3. glad, fornøjet, tilfreds, (himmel)henrykt, eksalteret, jublende

De har sjældent samme denotation, og det skal de heller ikke; de skal inspirere den skrivende, inden han har lagt sig fast på et præcist indhold. *Glad* og *fornøjet* har fx ikke samme denotation, eftersom *glad* spænder over større glæder, end *fornøjet* kan leve op til. (*Moderen blev glad, da hun genfandt sin 3-årige datter* – forhåbentlig ikke blot *fornøjet*).

Ved ÆGTE SYNONYMER forstår jeg udtryk med samme denotation, fx:

4. navneord–substantiv, omkreds–periferi, sælhund–sæl, århundrede–sekel, trækbasun–trombone

Synonymikken gælder naturligvis kun inden for en bestemt fælles betydningsvariant. Ordene *hest* og *springer* er ægte synonymmer, men kun i deres betydningsvariant vedr. skak, ligesom eksemplet med *omkreds* kun vedrører geometri-varianten. Ord, der i samtlige deres betydningsvarianter er ægte synonyme, er ligeså interessante som ord, der staves ens forfra og bagfra.

Ved MAKSIMALT SYNONYME udtryk forstår jeg udtryk med samme denotation og konnotation. Her må det straks slås fast, at to forskellige stimuli, som man klart kan høre forskel på, *logisk tvingende* har forskellig fysisk og psykisk virkning, fx æstetisk virkning. Desuden siger det sig selv, at ordenes associationer delvis er individuelle. I rekordernes land, USA, blev der i mange år holdt skønhedskonkurrence mellem sprogets ord, og det var hvert år ordet *mother*, der vandt. Men der *må* jo være amerikanere, der har et belastet forhold til dette ord! Det giver derfor kun mening at tale om maksimalt synonyme ord inden for den enkelte idiolekt. Alene dette, at sprogbrugernes ordforråd er vidt forskellige, tilsiger det. Hvis MAKSIMAL SYNONYMI skal være en brugbar, sprogligt relevant klasse – og ikke en tom og selvmodsigende klasse – må man forlade det utopiske (at alle oplever samme konnotationer) og stille sig tilfreds med, at man i én og samme idiolekt ikke

umiddelbart kan pege på nogen ladnings-, stil-, associations- eller anden konnotativ forskel. I min idiolekt er flg. ord max. synonyme:

5. trækkrog–trækkekrog (forskellige ord!), Køgebo–Køgeboer, dippe-dut–duppedit, bøg–bøgetræ, sproglære–grammatik, sludder–vrøvl, anholde–arrestere.

Sproglære er i dag så sjældent, at det for yngre sikkert er afstikkende, perifert. Men jeg har læst og hørt det så tit, at det umiddelbart fremtræder helt naturligt og upåfaldende. Jeg ved også, at politiet siger *anholde*, men for mig (og andre) er *arrestere* ligeså normalt.

3. Stikprøver

I det flg. ser vi på nogle ordbøger, hvori jeg har udvalgt nogle tilfældige artikler med mere end en (af redaktøren hævdet) ækvivalent inden for en enkelt betydningsvariant. De er helt mekanisk valgt. Jeg har modstået fristelsen til at slå ned på ekstra store tag-selv-borde. Gyldendals røde dansk-tyske og tysk-danske (begge Bork m.fl. 1999), har:

6. dybbå ~ tiefblau, sattblau.

Principielt er jeg imod *sattblau*: Den aktive danske amatør, der skal tale/skrive tysk, har ingen brug for det; han er altid på den sikre side med *tiefblau* og har ikke brug for stilistisk variation. Han vil desuden være usikker på, om *sattblau* nu også betyder nøjagtig det samme. Den passive tyske bruger, der skal *forstå* en dansk tekst, får intet ud af at blive belært om de to muligheder i sit eget sprog. Kun den professionelle skønlitterære oversætter til tysk (i praksis en tysker) kunne tænkes at blive glad for den stilistiske variationsmulighed, som han dog typisk let ville kunne finde selv. Men det har jo en pris! Artiklen kræver lidt større opmærksomhed, og længere tid at tilegne sig, og den bliver betydelig sværere at huske.

7. adresse ~ 1. (*bopæl, påskrift på brev*) Adresse, Anschrift; 2. ...

Jeg fornemmer, at *Anschrift* er mere formelt end det helt neutrale *Adresse*. Dermed er det forkert, når der ikke står "(mere formelt) Anschrift", men selv

med en sådan kommentar hører det ikke hjemme ved det stilistisk neutrale *adresse*. Men igen: Selvom dette synonym var perfekt, ville ingen have glæde af det, hverken skoleeleven, den lystlæsende tysker eller den skønlitterære oversætter.

8. kysse ~ küssen (fx er küsste sie auf den Mund; er küsste ihr die Hand); (*grundigt*) ab-küssen; ...

Der er fejl på fejl i den sidste oplysning. *Abküssen* betyder slet ikke 'kysse grundigt'. If. den ty.-da. søsterordbog betyder det *enten* 'kysse hjerteligt' (fx om mand og kone) *eller* nedsættende 'over(be)kysse'. Men dette underbegreb til *kysse* hører slet ikke hjemme her. For *kysse* har aldrig denne betydning. Skulle den skønlitterære tyske oversætter have glæde heraf den dag, han møder en tekst a la *Hun kyskede ham hjerteligt?* Måske, men prisen er for høj. Artiklerne bliver ulæselige med sådanne underbegreber, der aldrig er påkrævede.

9. bryllup ~ Hochzeit, Vermählung; ...

De to synonymer er vistnok ægte, men jo på højst forskelligt stilniveau. Synonymitis.

10. Eingeständnis ~ indrømmelse, vedgåelse, tilståelse.

Alle tre synonymer kan bruges, inflatorisk, om ikke-fejltrin (*Han indrømmede, at han blev forpustet – Jeg indrømmer, at jeg var i Hobro den dag, men ikke at jeg...*), og alle tre kan bruges udenfor og indenfor en politimæssig sammenhæng. På dansk er *tilståelse* det "alvorligste" af dem, det, der bedst passer på regulære moralske fejl og dermed bedst under forhør. Den bedste ækvivalent til *tilståelse* er vistnok *Geständnis*, sådan at *Eingeständnis* passer bedst til *indrømmelse*. (At vurdere dette, som udlænding, på net og i ordbøger er meget svært; bedre er vor leksikografi desværre ikke). Men selvom jeg skulle tage fejl, og de alle passer lige godt på *Eingeständnis*, hvem har så glæde af tripletten? Ingen – end ikke den danske skønlitterære oversætter (som næppe slår sprogets 10.000 almindeligste ord op). Trænger han til variations-inspiration, slår han op i en synonymordbog. Den læge bruger bliver usikker og vil ikke kunne huske artiklen ret længe.

Gyldendals store røde engelsk-danske (Kjærulff Nielsen m.fl.1995) har:

11. beam [bi:m] *vb.* 1. støtte (, forsyne) med bjælker; 2. ...

At man i stedet for *huset skulle støttes med bjælker* kan sige *huset skulle forsynes med bjælker* hjælper ingen og forvirrer mildest talt.

12. conveyance *sb.* 1. a) transport; kørsel; befordring; ...

Trods semikolonerne er der tale om én eng. betydningsvariant, som dækkes fint af *befordring*, som er et formelt ord ligesom *conveyance*. At *transport* og *kørsel* har betydningsvarianter med samme denotation, men ikke samme stil, kan kun forvirre i denne højst kompakte artikel (her forkortet). Og skulde en aktiv bruger i en bestemt kontekst finde *befordring* uegnet, er det let nok, i ens eget sprog, at finde alternativer: *kørsel*, *udkørsel*, *spedition*, *varetransport*. En ordbog kan alligevel aldrig udtømme alle konkrete løsninger.

13. lubber *sb.* ... 1. (*clumsy person*) klods, klodrian. 2. ...

Synonymerne er ægte og har samme stil, kun metaforikken skiller en anelse. *Klodrian* er ikke metaforisk; *klods* minder måske, langt ude, visse tankefulde om et stykke legetøj. Men igen er det kun den professionelle oversætter, der måske bliver glad for det overflødige *klods*, fordi han ikke lige tænkte på det ord. Men prisen for de hundreder af sagprosa-brugere og passiv-brugere er for høj. Gyldendals røde fransk-danske (Sørensen m.fl. 1978), skriver:

14. horizon ... *m.* synskreds, horisont (også *fig*).

Ægte synonymer, men den, der gengiver det neutrale *horizon* med det udpræget litterære *synskreds* (som højst 2 % af danskerne kender), oversætter galt. – I den nyeste udgave (Sørensen m.fl. 2014) er rækkefølgen af de to byttet om. Lapperi.

4. Hvornår er dubletter berettiget?

Det er de først og fremmest, hvor udgangs-ordet klart savner en ækvivalent, men dækkes af et samlet felt af flere ord i målsproget – som vi så det med

sv. *stilig*. Eller i ty. *Großmutter* ~ sv. *farmor/mormor*, da. *by* ~ eng. *city/town/village*. Se Brink 2008 om leksikalske ”huller”.

Et andet tilfælde er, hvor synonym nr. to mest tjener til at forklare nr. et – enten fordi nr. et er et ret ukendt ord, eller fordi nr. to skal præcisere nr. et’s betydningsvariant. Fx:

15. *svensk* egid ~ *dansk* ægide (beskyttelse; ledelse)

Eksemplet er fra den glimrende nye svensk-danske ordbog (Kristensen m.fl. 2010), der, yderst tilfredsstillende, typografisk markerer, at parenteser ikke rummer ækvivalenter, men forklarer ækvivalenten. Endvidere (fra Kjærulff Nielsen m.fl.1995):

16. *eng.* station ~ *da.* 1. station, jernbanestation. 2. ...

Den korrekte ækvivalent er *station*, men det er pinende nødvendigt at vise, hvilken betydningsvariant det optræder i. Jeg foretrækker forklarende parentes: 1. (*about railways*) station. 2. ...

5. Hvordan er miseren opstået?

Hovedårsagen til synonymitis er, tror jeg, at leksikografen ser sin situation som den, at han som fagmand bliver spurgt af en anden oplyst landsmand om, hvad dét og dét hedder på et andet sprog. Hvis nogen fx spurgte mig om, hvad fx et lufthul hedder på spansk, ville jeg elske at udbrede mig med: *ba-che de aire*, og straks tilføje, at det ligeså vel kan hedde *pozo de aire* og *bolsa de aire*, ligesom jeg ville nyde at gennemgå de næsten umærkelige metaforik-forskelle og dermed konnotations-forskelle. Jeg kunne også finde på, at meddele ikke-ækvivalenter som en service. Fx sige om dansk *opmærksom* (høflig), at det hedder *obsequioso*, men at spørgeren måske også kunne have glæde af overbegreberne *cortés* og *galante* (begge ’høflig’). Sådan gør den store dansk-spanske ordbog (Hansen & Gawinsky 2004) – uden at kommentere ”ækvivalenterne”. Men ordbogsbrugeren er i en helt anden, skolestuefjern situation: Den L1-aktive skal bruge én ækvivalent, den bedste, for han taler ikke som Else Skolemesters i synonymmer. Den L2-passive og -aktive ønsker at forstå, de skal ikke belæres om deres eget modersmål.

6. Konklusion

Man må erkende forskellen mellem *ækvivalens* og *oversættelse* (som ikke er ny). Eng. *tulip* har én eneste ækvivalent på dansk: *tulipan*. Men i en konkret tekst om tulipaner kan man da udmærket i stedet for *the tulip* sige *blomsten*. Der hverken tilføjes eller mistes nogen information. Man kan også, hvis det altså handler om tulipaner, sige *Hun foretrak de gule* i stedet for. ... *the yellow tulips*. Men sådanne muligheder kan slet ikke udtømmes, og skal det heller ikke. Ordbogen skal give *ækvivalenten*, dvs. det, som glosen i sig selv signalerer: denotation og konnotation. Så kan aktivbrugeren improvisere herover, hvis han har lyst.

Hvis ordbogen angiver to (næsten) maksimalt synonyme ord, fx *lastbil*, *lastvogn*, begår den jo ingen fejl, men den komplicerer artiklen og gør den sværere at huske – til glæde højst for den skønlitterære oversætter. Skoleeleven, forretningsmanden, journalisten, lystlæseren har, passivt såvel som aktivt, ikke brug for begge synonymer. Den professionelle oversætter må henvises til en synonymordbog eller sin egen fantasi vedr. sit eget sprog, tungemål, signalsystem ...

Litteratur

Ordbøger

- Bork, Egon m.fl. (1999): *Dansk-Tysk Ordbog*. 11. udg. København: Gyldendals røde serie.
- Bork, Egon m.fl. (1999): *Tysk-Dansk Ordbog*. 14. udg. København: Gyldendals røde serie.
- Hansen, Johan Windfeld & Birthe Gawinski (2004): *Dansk-spansk ordbog*. 5. udg. København: Gyldendal.
- Kjærulff Nielsen, B. m.fl. (1995): *Engelsk-Dansk Ordbog*. 5. udg. København: Gyldendal.
- Kristensen, Kjeld m.fl. (2010): *Svensk-Dansk Ordbog*. København: DSL.
- Munch-Petersen, Valfrid Palmgren m.fl. (1966): *Svensk-dansk Ordbog*. 2. udg. København: Gyldendal.

- Sørensen, N. Chr. m.fl. (1978): *Fransk-dansk Ordbog*. 7. udg. København: Gyldendal.
- Sørensen, N. Chr. m.fl. (2014): *Fransk-dansk Ordbog*. 10. udg. København: Gyldendal.

Anden litteratur

- Brink, Lars (2008): Leksikalske huller. I: Ásta Svavarsdóttir, Guðrún Kvaran, Gunnlaugur Ingólfsson & Jón Hilmar Jónsson (red.): *Nordiske Studier i Leksikografi* 9. Reykjavik: Nordisk Forening for Leksikografi, 61-67.
- Brink, Lars (2010): Pålydende-betydning i ordbøger. I: Harry Lönnroth & Kristina Nikula (red.): *Nordiska studier i lexikografi* 10. Tammerfors: Nordisk Forening for Leksikografi, 113-117.
- Schibbye, Knud (1969): *Engelsk synonymik på grammatisk grundlag*. København: Nordisk Sprog- og Kulturforlag.

Lars Brink
professor i dansk sprog og litteratur
Elmevænget 30
DK-2880 Bagsværd
labri@webspeed.dk

Inventariografi og andre nyere tilgange til leksikografi

Ken Farø

This article presents an analysis of four recent approaches to lexicography, whose common denominator is *method radicalism*, namely (1) The Corpus Method, (2) The Function Doctrine, (3) User-driven lexicography, and (4) Inventoriography. An inventoriographic approach to list-oriented language description, where language lexicography represents a significant, but not sufficient pillar, hails the slogan “Back to language”, i.e. a type of lexicography that to a larger extent than the existing focuses on the description of individual linguistic categories and their complete inventory. A few examples of inventories are the prepositions, text genres, names, book titles, phonemes, templates, discourses etc. – in other words: a great task. It is an intrinsic part of Inventoriography to avoid artificial boundaries between lexicography, encyclopediography and grammaticography.

1. Indledning

Hvorfor kan man ikke

- kombinere en semasiologisk ordbog med et systematisk register?
- søge direkte på forskellige delinventarer i elektroniske ordbøger?
- kombinere en grammatik med en ordbog og et leksikon?
- udvide hhv. leksiko-, grammatiko-¹ og encyklopædiografiens² genstands-områder?
- beskrive grammatiske elementer lige så systematisk og tilgængeligt som leksikalske og encyklopædiske?

Svar: Det kan man også, men det er ikke gængs praksis. Vi skal i det følgende se eksempler på, hvordan det kan gøres, og hvad det indebærer.

¹ = grammatikskrivningen.

² = leksikonskrivningen.

Dette bidrag er samtidig en analyse af fire relativt aktuelle tilgange til leksikografi, hvis fællesnævner er metoderadikalisme³ (jf. Farø 2006), nemlig (1) Korpusmetoden (Asmussen 2004), (2) Funktionslæren (Tarp 2009), (3) Den brugerefterspørgselsstyrede leksikografi (BEBL), og (4) Inventariografien (Farø 2013). Det er den sidste tilgang, bidraget selv hviler på.

2. Korpusmetoden

Korpuslingvistik er efterhånden en velkendt og velafprøvet metode til at opbygge opslagsværker ud fra. Eksempler på i høj⁴ grad korpusbaserede publikationer er bl.a. MacMillan English Dictionary (2013, = MED) og Den Danske Ordbog (2003-05, = DDO). I en korpuslingvistisk ordbog af moderne tilsnit bestemmer korpusindretningen + den valgte filtreringsgrad den konkrete udvælgelse af opslagsmateriale. Med korpusmetoden opnår man en bedre og også mere relevant dækning af ordforrådet. Man får desuden en langt større præcision i sprogbeskrivelsen, fordi computere er i stand til at opfange det, som ikke er særlig påfaldende for den menneskelige hjerne. Mennesker er relativt dårlige til eksplicit at analysere det almindelige og hyppige (se også Farø 2015a). Man opnår herudover mere dybde i analysen, herunder fx valensbeskrivelse, kollokationsanalyse, fraseologi (jf. Farø 2015b og *LexicoNordica* bind 16) og en meget differentieret semantisk inddeling. Endelig giver metoden mere (intenderet) ensartethed i artikelopbygningen, fordi skriveprocessen er nemmere at overvåge elektronisk end manuelt. Det er bl.a. takket være korpusmetoden, at den nyere danske interjektion *okay* bliver beskrevet med 4-5 betydningsnuancer i DDO, en funktionel differentiering de færreste sprogbrugere er bevidste om.

Det er i den monolingvale leksikografi, vi finder korpustilgangen, ikke i den bilingvale. Det er ikke overraskende, men det er på den anden side heller

³ Hermed mener jeg en konsekvent tilrettelæggelse af opslagsgenstanden ud fra én metode.

⁴ I hvert fald DDO er ikke 100 % korpusbaseret; og det er helt bevidst. Selv om nogle vil finde konceptionen ærgerligt mudret, blev det sent i processen besluttet, at ordbogens nytteværdi ikke skulle mindskes af korpussets eventuelle tilfældige mangler. I det omfang redaktørerne efterfølgende konstaterede åbenlyse lakuner i ordbogen, blev de udbedret af hensyn til brugerne. Det videnskabelige eksperiment blev på dette punkt prioriteret lavere end bogens praktiske værdi for brugerne.

ingen naturlov, at tosprogsordbøger i højere grad er ikke-empirisk baserede. Det er et oplagt fremtidsmål at producere bilingvale ordbøger, der ikke bare er korpusbaserede og dermed langt mere pålidelige og præcise end nu, men som også indeholder oplysninger om fx frekvensforhold, evt. tidligere misforståelser (leksikografisk historik), ækvivalensproblemer (pragmatisk, tekstvidenskabeligt, stilistisk, sociolingvistisk, diakront m.m.). Pladsen behøver man ikke bekymre sig så meget om, for det er naturligvis inden for rammerne af e-leksikografi, vi taler. Og her handler det ikke om plads som sådan, men om hvad der skal stå i for- eller baggrunden.

I *Gyldendals Røde Ordbøger tysk-dansk* (netversion) får man under opslaget *Sie* meget ukompliceret det danske høflighedspronomen *De* som ækvivalentforslag. En grundig korpusundersøgelse afdækker imidlertid et andet forhold: 1. *Sie* er relativt langt hyppigere end *De*; 2. Det danske *du* er omvendt relativt meget hyppigere end det tyske *du*. Hvorfor? Fordi det tyske *Sie* er den umarkerede form, mens det danske *du* i sit system er umarkeret.

Korpusanalysen vil desuden kunne afdække, hvordan de 2 x 2 pronomen anvendes hver især og dermed kunne vejlede brugeren pragmatisk bedre, end det er tilfældet nu. Fx kunne man forestille sig flg. tekst i en dansk-tysk ordbog:

du:

1. (ved ukendt tiltalt voksen, i professionelle og offentlige situationer samt ved aldersforskel): *Sie*
2. (privat, familiært, blandt venner og yngre jævnaldrende): *du*.

3. Funktionslæren

En anden nyere tilgang til leksikografi er den især århusiansk funderede Funktionsleksikografi, som udviklede sig stærkt i 90'erne og 00'erne. Bag paradigmet står først og fremmest forskerne Bergenholtz, Tarp, Leroyer og Nielsen, der både har leveret en lang række teoretiserende, ofte stærkt kritiske og præskriptive diskussionsbidrag, samtidig med at gruppen selv, med forskellige samarbejdspartnere, har udviklet en hel del, ikke mindst faglige ordbøger.

Alle leksikografiske valg afhænger ifølge denne skole af, hvilke leksikografiske funktioner ordbogsværktøjet⁵ (!) skal løse for brugeren, hvem denne er, og i hvilke(n) brugssituation(e)r han/hun befinder sig. Ligesom korpus-tilgangen satte fokus på empirien og den optimerede analyse af den, har Funktionslæren tilført faget en lang række nye perspektiver på selve ordbogen og forholdet mellem brugerens behov og ordbogens indretning.

Funktionslæren svarer lidt til, hvad Vandermeer med sin skoposteori forlangte af oversættelsen: Hele arbejdet med ordbogens indsamling og indretning er i lige linje et produkt af, hvad den overordnet set skal løse af behov og for hvem. Uden denne (grundige) indledende status giver hele projektet ingen mening ifølge Århus-skolen. Og alle inventarrelaterede spørgsmål er ligeledes en relation af de leksikografiske funktioner.

Det er som nævnt en væsentlig pointe, men i praksis har den vist sig at være lidt for sejrssikker på en måde, der i en række tilfælde har ført til, at selve den sproglige materie ikke altid har været behandlet optimalt, fordi fokus synes at have været på metaleksikografi og i mindre grad på det sproglige inventar og deskriptiv metode (se fx Farø 2005).

4. Brugerefterspørgselsbaseret leksikografi (BEBL)

I de senere år har der de facto udviklet sig en leksikografisk tilgang, som udmærker sig ved at være særdeles arbejdsintensiv og, i hvert fald på kort sigt, økonomisk rationel. Dens kerneide er, at ordbogens indhold skal være defineret af den efterspørgsel, brugerens søgeadfærd afslører i form af spor i den tilhørende logfil, som er knyttet til søgefunktionerne på den hjemmeside, der *hoster* ordbogen.

Den brugerefterspørgselsorienterede tilgang er ikke egentlig teori- eller metodeorienteret. Den er først og fremmest økonomisk motiveret ud fra devisen: "This is what you want, this is what you get!". Tilgangen forsøger primært at opfylde den enkelte kundes efterspørgsel, enten eksplicit (direkte via brugerens forespørgsler) eller implicit (indirekte ud fra søgeadfærden).

Den økonomisk orienterede tilgang til genstanden medfører et lavt ambitionsniveau med hensyn til research, sprogbeskrivelse, fuldstændighed og

⁵ Ordbøger er ifølge Funktionslæren *altid* værktøjer, og ikke andet. Dette synspunkt deler den foreliggende tilgang ikke fuldstændigt (se nedenfor).

systematik. Derfor erstatter BEBL ikke en videnskabelig deskriptiv leksikografi ideelt set, men måske desværre nok i praksis.

Den afgørende forskel er vægtningen mellem systematisk og punktuelt redigering. BEBL står for punktuelt opbygning og redigering.

I Danmark er dette felt frem for alt domineret af Ordbogen.com og Gyldendals online-ordbøger, men det fylder også en del i andre internetordbøger.

Den brugeradfærdsstyrede leksikografi leverer pr. efterspørgsel. Det fører til øget heterogenitet og asymmetri i beskrivelsen af ordbogsmaterialet, noget der ikke spiller nogen vigtig rolle for selvsamme tilgang. For den stiller heller ikke andet i udsigt end at forsøge at give svar på den enkelte brugers spørgsmål. Hvorvidt den gør det tilfredsstillende eller snarere: om kvaliteten af de produkter, der baseres på baggrund af eksterne brugeres søgninger foretaget i denne type opslagsværker, er acceptabel, kan ikke afgøres her.

Fra min erfaring på et kommercielt⁶ forlag ved jeg, at den BEBL i nogle tilfælde fører til meget specifikke, lokale og faglige lemmata, fordi fokus ikke længere er på den overordnede ordbog, strukturer, ja på leksikografi, men snarere på at imødekomme den singulære kundes aktuelle efterspørgsel. I virkeligheden er dette ikke kun en potentiel risiko for den moderne kommercielle leksikografi, men mere generelt for meget brugernære og brugerkommunikerende ordbøger.

5. Grammatikografi, leksikografi og encyklopædiografi

Mens leksikografien i hvert fald siden 1980'erne (jf. bl.a. Hausmann et al. 1989-91) har befundet sig i en voldsom udvikling, både teoretisk, metodisk og teknologisk, har der i grammatikskrivningen været mere konceptuel stilstand. Man kan sige, at der generelt hersker et lavt metagrammatikografisk teoriniveau. Formidlingsrammerne er ikke i fokus her. Modsat forventningen er mange grammatikker præget af meget brødtekst snarere end en tilgangsvenlig struktur (eksempler på dette er fx Eisenberg 2006 og Lauridsen & Poulsen 1995). Her kunne grammatikografien godt lære af leksikografien, der er formidlingsmæssigt mere selvrefleksiv og dynamisk.

⁶ Brugt i ordets helt neutrale betydning.

De tre angiveligt distinkte områder grammatik, ordforråd og encyklopædi er i praksis ofte konkret forbundne i bl.a. tekster, indlæringsammenhæng og oversættelse. Ved både skriftlig og mundtlig reception af løbende tekst forekommer grammatiske, leksikalske, encyklopædiske udfordringer m.v. ind imellem hinanden. En bruger er derfor ikke altid hjulpet med en rent leksikalsk opslagsressource. Løsningen på disse problemer ligger for mig at se i den tilgang, jeg nu vil behandle.

6. Inventariografi

Med konceptet *inventariografi* etableres den overbygning, som både grammatikografien og leksikografien logisk mangler. En inventariografisk tilgang til opslags- og oversigtsorienteret sprogbeskrivelse, hvoraf (sprog)leksikografien udgør en væsentlig, men ikke uantastelig søjle, hylder slagordet ”Tilbage til sproget!”.

Hermed mener jeg en leksikografi, der i højere grad end den eksisterende fokuserer på beskrivelse af de enkelte sproglige kategorier og deres inventar. Fx præpositionerne, tekstgenrerne, navnene, de bevingede ord, bogtitlerne, fonemerne, skabelonerne, diskurserne, matematikkens sprog⁷ (jf. Jagers 2009) osv., alt sammen under hensyntagen til deres respektive underopdeling. Gerne med hjælp fra korpuslingvistikken og Funktionslærens refleksionsapparat, mens Efterspørgselsprincippet overlades til de kommercielle redaktioner, hvor det giver god mening, og på ingen måde behøver at komme i klammeri med Inventariografien. Men netop med dette slagord står Inventariografien principielt i modsætning til både Funktionslæren og Efterspørgselsprincippet, mens den er lidt mere, men ikke fuldstændigt kompatibel med korpusmetoden.

Inventariografien forfægter den systematiske sprogbeskrivelse, hvormed den tilsyneladende er traditionalistisk. Intet kunne imidlertid være mere forkert: Inventariografien lægger nemlig særlig vægt på at undgå oprettholdelse af kunstige grænser mellem leksikografi og grammatikografi –

⁷ Artiklen taler om ”Matematikens ord”, men et fagligt domæne omfatter typisk også særlige konstruktioner, orddannelseselementer, fraser og tekstgenrer, hvilket en inventariografisk tilgang ville tage højde for.

hvormed vi i øvrigt slet ikke har udtømt det sproginventariografiske beskrivelsespotentiale.

Inventariografien er en metadisciplin for al relativt ensartet systematisk, overbliksskabende og tilgangsorienteret sprogbeskrivelse. Det vil sige, den er tænkt som moderdisciplin for både leksikografi, grammatikografi og encyklopædiografi. Men samtidig er den et samlebækket for de sproglige inventarer, der typisk *ikke* omfattes af et af disse tre områder, nemlig fx navne af enhver slags, tekstgenrer, talehandlinger, bevingede ord, etc. Det vil sige alle former for *-grafier*.

Hvor Funktionslæren kun anerkender ordbøger (og vel også grammatikker og leksika), der er udarbejdet med bestemte brugsfunktioner for øje, rummer en inventariografisk tilgang mulighed for sprogkompilation og -deskription alene for *dokumentationens* skyld.

6.1. Typer af inventarer

Med termen ”inventar” mener jeg både sprogets samlede mængde tegnklasser OG de enkelte klassers inventarer. Fx modalverber, toponymer, interjektioner, fleksiver og tekstgenrer. De specifikke inventarer kan inddeles ud fra den klassiske skelnen mellem hhv. lukkede (fx præpositioner, artikler, fleksiver, fonemer, affikser, ordsprog) og åbne (fx ord, frasemer, navne, titler, konstruktioner, skabeloner, tekstgenrer).

Inventarisering kan finde sted på adskillige måder: Man kan nøjes med kompilationen og systematiseringen, fx i en database, som så kan danne grundlaget for undersøgelser. Eller man kan anvende materialet som kernen i en publikation, en opslagsressource.

Nedenfor ses en oversigt over væsentlige respektive træk ved de 4 tilgange til systematisk, liste- og kategoriorienteret sprogbeskrivelse:

	Korpus- leksikografi	Funktions- leksikografi	Bruger- efterspørgsels- leksikografi	Inventariografi
Vigtigste kompila- tions- kriterium	sprogbrugs- frekvens	bruger- og brugsrelevans (abstrakt)	brugernes direkte forespørgsler, analyse af brugeradfærd	alle typer sproglige inventarer, fuldstændighed, dokumentation
Styrker	beskrivelses- præcision, autenticitet	stærk generel brugerrelevans	stærk individuel brugerrelevans	helhedsorien- tering, linkning
Svagheder	vilkårlige medtagelses- kriterier	sproget i sig selv relativt uinteressant	sprogbe- skrivelse uden selvstændig værdi	langtidsprojekt, kun egnet til e- publicering

7. Ideel arbejdsdeling mellem de fire tilgange

Som det allerede vil være fremgået, er de fire nyere tilgange til listeorienteret sprogbeskrivelse ikke nødvendigvis komplementære: Man kan i stedet for at anskue dem som absolutte principper betragte dem som metoder, der supplerer hinanden, fx på følgende måde:

1. Opslagsressourcer udtænkes grundlæggende funktionsadækvat og inventariografisk.
2. Grammatiske og leksikalske elementer beskrives i sammenhæng og linkes til hinanden.
3. Korpusteknologi og -metode anvendes som datagenerator i samspil med Funktionslæren, suppleret med inventariografiske tjek.
4. Bevidsthed om mangler opnået via Brugerefterspørgselsmetoden anvendes som supplement til den redaktionelle vedligeholdelse af ressourcen.

8. Sammenfatning

Vi mangler et begreb og også en konkret disciplin, der forener ikke bare leksikografi og grammatikografi, men også alt det i sproget, der typisk ikke

behandles af ordbøger og grammatikker. *Inventar* og *inventariografi* er præcise og dækkende termer til formålet, mens *ordbog* og *leksikografi* begge i mange tilfælde ligger langt fra det, de faktisk indeholder. Betegnelsen *leksikografi* spærrer i nogle tilfælde for indsigten, at det i en bestemt sammenhæng måske ikke ”bare” er en ordliste, der er brug for, fx når det gælder om at afdække en bestemt forfatters eller et bestemt domænes sprogbrug. Med den meget bredere disciplin *inventariografi* kunne vi fx i dette forum begynde at hjælpe grammatikerne med at lave lettilgængelige, overskuelige, og søgbare grammatikker i stedet for de ofte meget fortællende og svært navigerbare grammatikker, vi stadig ser. Grammatikker udgives i højere grad end ordbøger på tryk, men det skyldes snarere, at forfatterne primært er interesserede i grammatik og tilsyneladende ikke i samme grad i formidlingen og indlæringen af den. Vi har brug for gode inventariseringer af både modersmål og fremmedsprog til brug for bl.a. didaktiske, receptive, oversættelsesmæssige og dokumentariske formål, herunder sproghistoriske.

Sprogene skal med tiden inventariseres, også interlingvalt. Det er den ultimative beskrivelsesopgave set fra et lingvistisk synspunkt. At den er omfattende (jf. Lorentzen 2015: 16), skal ikke afholde os fra at tage udfordringen op.

Litteratur

Ordbøger

DDO = Den Danske Ordbog. <www.ordnet.dk/ddo> (januar 2016).

DUW = Duden Universalwörterbuch (2015, App).

GRO = Gyldendals Røde Ordbøger (netversion). <legacy-ordbog.gyldendal.dk> (januar 2016).

MED (2013) = MacMillan English Dictionary. London: MacMillan Education Limited.

Ordbogen.com. <www.ordbogen.com> (januar 2016).

Anden litteratur

- Asmussen, Jørg (2004): Korpus 2000 – til hvilken nytte? Muligheder og grænser for empiriske sprogundersøgelser. I: Duncker, D. (red.): *Studier i Nordisk 2002-2003*. København: Selskab for Nordisk Filologi, 17-19.
- Eisenberg, Peter (2006): *Grundriss der deutschen Grammatik: Das Wort*. Stuttgart: Metzler.
- Farø, Ken (2006): Leksikografisk radikalisme? 10 forslag til den bilingvale leksikografi. I: Lorentzen, H. & L. Trap Jensen (red.): *Nordiske Studier i Leksikografi* 8. København: Nordisk Forening for leksikografi, 85-96.
- Farø, Ken (2013): Inventariografi – al leksikografis og grammatikografis mo(r)der? I: *LexicoNordica* 20, 35-55.
- Farø, Ken (2015a): Gak til korpus og bliv viis: om korpora, sprognormer og tysk. I: Duncker, Dorte, Ole Ravnholt & Eva Skaft Jensen (red.): *Rette ord – Festskrift til Sabine Kirchmeier-Andersen i anledning af 60-årsdagen*. Dansk Sprognævns skrifter 46. København: Dansk Sprognævn, 109-123.
- Farø, Ken (2015b): Feste Wortgruppen/Phraseologie II: Phraseme. I: Haß, U. & P. Storzjohann (eds.): *Handbuch Wort und Wortschatz*. Handbücher Sprachwissen, vol. 3. Berlin/Boston: De Gruyter, 226-247.
- Farø, Ken & Henrik Lorentzen (2009): De oversete og mishandlede ordforbindelser. I: *LexicoNordica* 16, 75-101.
- Hausmann, Franz Josef, Oskar Reichmann, Herbert Ernst Wiegand & Ladislav Zgusta (Hrsg.) (1990): *Wörterbücher. Ein internationales Handbuch zur Lexikographie* 2. Berlin/New York: De Gruyter, 2037-2049.
- Jagers, Peter (2009): Matematikens ord. I: *LexicoNordica* 16, 315-318.
- Lauridsen, Ole & Sven-Olaf Poulsen (1995): *Tysk Grammatik*. København: Munksgaard.
- LexicoNordica* 16 (2009) (temanummer om fraseologi). Oslo: Nordisk forening for leksikografi.
- Lorentzen, Henrik (2015): Da NFL kom til København. I: *Ledant* 60, 14-16.
- Tarp, Sven (2008): *Lexicography in the Borderland between Knowledge and Non-Knowledge, General Lexicographical Theory with Particular Focus on Learner's Lexicography*. Tübingen: Niemeyer.

Ken Farø
lektor, ph.d. i tysk sprog
Institut for Engelsk, Germansk og Romansk
Københavns Universitet
Njalsgade 128
DK-2300 København S
kenfaroe@hum.ku.dk

Hvordan slippe inn i Platons hage¹? – Om kartlegging og dokumentasjon av norsk akademisk vokabular

Ruth Vatvedt Fjeld & Arash Saidi

In the last few years several dictionaries of academic words have been developed, especially for English, but also for other languages, based on automatic analysis of academic corpora. Different methods have been used. For the Norwegian Academic Vocabulary list, we have experimented with different types of stop lists, and the article presents some results and a discussion of the lexical results of these experiments. We also recommend a phraseological analysis of academic corpora to map out some of the typical phrases in academic language.

1. Hvorfor er akademiske ordlister viktig?

Manglende kjennskap til akademisk vokabular blir ofte regnet som årsak til store forskjeller mellom studenters akademiske prestasjoner. Både første- og andrespråksbrukere i alle utdannings situasjoner klarer seg bedre dersom de behersker et større og mer akademisk ordforråd:

Control of academic vocabulary, or the lack thereof, may be the single most important discriminator in the ‘gate-keeping’ tests of education. (Gardner & Davies 2013:1)

Det gjelder også for norske studenter. Dessuten leser de mye faglitteratur på engelsk, og engelsk blir ofte krevd i akademisk skiving. Dermed mangler studentene gjerne norske ekvivalenter og får probmeler når de skal skrive på morsmålet. Pedagogiske lister over norsk akademisk vokabular kan bøte på dette. Målet er å utvikle norsk til et fullverdig moderne akademisk språk.

¹Ordet akademi stammer fra Platons skole for filosofi og læring i Athen, grunnlagt ca. 385 f.Kr. Skolen lå i nærheten av et tempel i utkanten av byen, som etter tradisjonen skulle være oppkalt etter helten Akademos. (Kilde: Wikipedia)

Vi har få undersøkelser av ordforråd eller forståelse av ord i norsk. Mest kjent er en studie av 80 ord i nyhetsspråk (Vinje & Østby 1981). I undersøkelser av ordforståelse testes forståelse av vanskelige ord, sjelden hvordan de blir brukt i utforming av tekster. For akademisk framgang er produksjon av akademisk tekst like viktig som resepsjon. I tillegg til å kartlegge vanskelige ord vil finne ut hvilke ord og formuleringer som brukes for å organisere og strukturere tekst etter akademiske prinsipper. Det kan være alminnelige ord med spesiell funksjon i akademisk uttryksmåte.

I arbeidet med å kartlegge norsk akademisk vokabular følger vi opplegg fra liknende prosjekter, med systematiske analyser av ordforrådet i allmennspråk, i akademisk språk og i teknisk fagspråk. En sammenlikning av ordbruken etter teksttype kan vise hva som kjennetegner det leksikalske inventaret i typisk akademisk tekst. Som materiale anvendes et elektronisk korpus av akademiske tekster der frekvens og spredning av ord analyseres. Ord som regnes som typiske akademiske, blir tatt med i den akademiske ordlisten. Resultatet skal utgjøre den første listen over norsk akademisk ordforråd, kalt The Norwegian Academic Vocabulary List (NAV).

2. Hva er akademisk vokabular?

Leksikografiens hovedoppgave er å dokumentere og beskrive ords egenskaper. Å dokumentere typiske ord i akademisk språk for å lage hjelpelister for akademisk opplæring er dels en pedagogisk, dels en leksikografisk oppgave. Problemet er å skille akademiske ord fra andre typer ord. Det har vært lange diskusjoner om skillet mellom fagspråk og allmennspråk, uten at man har funnet en klar definisjon på forskjellen, f.eks. i Molde (1976:6) og Engberg (1992:94). Det er nok en glidende overgang mellom allment og faglig språk. De fleste forklaringer tar utgangspunkt i sender/mottaker-modeller, altså hvem som skriver, og for hvem, mer enn språktrekk i tekstene. Et unntak er Ralph (1981), som mener at forskjellen er at ordforrådet er annerledes i fagtekster enn i allmenntekster på grunn av fagtermene. Målet vårt er å finne generelt og emneavhengig vokabular som skiller den akademiske diskursens vokabular fra allmennspråkets.

3. Tidligere kompilerte akademiske ordlister

Den mest kjente akademiske ordlisten er *Academic Word List* (AWL) (Coxhead 2000), men det fins mange flere, jf. Gardner & Davies (2013) for utfyllende oversikt. Ordlistene er laget for engelsk i pedagogisk hensikt, både for andrespråksbrukere og utrente studenter. For nordmenn egner listene seg som hjelpemidler i arbeid med engelske akademiske tekster, men det er klart behov for hjelpemidler også til slik tekstproduksjon på norsk.

I Norden arbeides det også med akademiske ordlister. Hittil fins *Svensk Akademisk Ordlista* (SAO) (Ribeck et al. 2013).

4. Norsk akademisk vokabular (NAV)

4.1. Kartlegging av akademiske enkeltord

I Oslo er nylig første versjon av en liste over akademisk vokabular i norsk bokmål ferdig. Prosjektet ble satt i gang ved Institutt for lingvistiske og nordiske studier ved Universitetet i Oslo av Janne Bondi Johannessen, Ruth Vatvedt Fjeld, Kristin Hagen og Arash Saidi. Innen LUNAS-nettverket arbeides det med å lage tekstdatabaser med søkefunksjonalitet som gjør det mulig å søke på tvers av dansk, norsk og svensk, som grunnlag for pedagogiske akademiske ordlister. I første omgang var målet å kompilere en liste av enkeltord som er typisk akademiske.

4.2. Materiale

Materialet er et akademisk korpus av masteroppgaver, phd-avhandlinger, tidsskriftsartikler og liknende fra DUO (Digitale utgivelser ved Universitetet i Oslo). Korpuset består av 9689 dokumenter og utgjør ca. 310 millioner løpeord. Av disse er ca. 100 millioner løpeord lemmatisert med Oslo-Bergen-taggen. Listen og de øvrige resultatene er derfor foreløpige, men vi regner med at de viser reelle tendenser i det norske akademiske bokmåls-vokabularet.

Tekstene kommer fra humaniora (44 mill. ord), pedagogikk (17 mill. ord), medisin (10 mill. ord), samfunnsfag (14 mill. ord), matematikk og

naturvitenskap (6 mill. ord), juss (4 mill. ord), teologi (3 mill. ord) og odontologi (4 mill. ord). Tekster fra alle fakulteter ved Universitetet i Oslo gjør at flest mulig akademiske stilarter ligger til grunn for analysene.

4.3. Metode for lemmaseleksjon

I tidligere akademiske ordlister har man unngått allmennspråklige ord, allmenne fremmedord og fagspesifikke, tekniske termer. Gardner & Davies (2013:8) framhever at typisk akademiske ord skal forekomme med en viss frekvens og spredning i flere typer akademiske tekster og i forskjellige akademiske disipliner, samt at de ikke er høyfrekvente i allmennspråklige tekster, for å bli tatt opp i Academic Vocabulary List (AVL).

Det anvendes bare negative kriterier, dvs. fravær av visse egenskaper ved ord eller uttrykk ved utvelgelsen av akademiske ord, det Gardner & Davies (2013:8) kaller akademiske kjerneord: “academic core words are those that appear in the vast majority of the various academic disciplines”. Det stilles altså krav til statistisk spredning over de forskjellige disiplinene og teksttypene. Med negative kriterier og heuristiske metoder kan resultatene bli tilfeldige. Og når lemmaseleksjonen er basert på statistiske metoder, kan resultatene variere avhengig av type og størrelse på korpus. Kravet om spredning i akademiske tekster er et forbedrende tilleggskriterium, men like fullt er det uklart hva et akademisk ord egentlig er.

En tidlig studie av akademisk ordforråd er utført av Martin (1976:92). Hans definisjon er positiv: “Academic vocabulary has in common a focus on research, analyses and evaluation - those activities which characterize academic work”. Martin regner dermed med at det ordforrådet som er best egnet til akademisk analyse og vurdering, kan regnes som akademisk. For å beskrive slik virksomhet kreves spesielle tekststrukturerende ord og uttrykk, ofte adverbialer, konjunksjoner og subjunksjoner i tillegg til de tekniske termene på fagfeltet.

Lemmutvalget i de presenterte akademiske ordlistene inneholder likevel mange alminnelige ord; AVLs ti viktigste er: *study, group, system, social, provide, however, research, level, result*, og SAOs ti første *dock, studie, beskriva, social, enligt, innebära, samt, form, betydelse, fall*. Det er vanskelig å se at de er typisk akademiske. Resultatene fra statistisk lemmaseleksjon må derfor diskuteres ut fra hva som er ”adekvat ordforråd”

for akademisk virksomhet (Ribeck et al 2013:370). Siden fagtermer er silt vekk fra korpus, vil resultatlistene med nødvendighet inneholde allmenne ord. Fordi de er hyppige i akademiske tekster, kalles de akademiske. Seleksjonen er da fundert på klassisk stilistisk metode (jf. Engberg 1998:37). En mer pragmatisk fundert tilnærming, der man konsentrerer seg om å undersøke hvilke språklige midler som brukes til å utføre visse faglige funksjoner (Engberg, *ibid.*), kan gi et annet resultat. I vårt tilfelle vil det være de funksjoner som kreves i akademisk argumentasjon.

Vi anvender den samme metoden som ble brukt i SAO, dvs. å kartlegge ord som ikke faller innenfor det allmenne og som ikke er fagterminologi. Det er likevel en viss forskjell mellom de to listene, da DUO-korpuset har spredning på flere fagfelt enn SweAk-korpuset (jf. Ribeck et al. 2013:375), som bare har tekster fra humaniora og samfunnsvitenskap. Som i SAO er det i NAV tatt utgangspunkt i lemmaer og ikke ordfamilier.

Vår metode var tredelt. Første steg var å finne et ords "keywordness", som definerer om et ord har uvanlig frekvens i korpus. Dette målet er beskrevet av Scott (1997:233-245), som framhever at ethvert ord med lavere "keywordness" enn 1.1 skal lukes ut av korpuset. Andre steg var å finne den reduserte frekvensen for hvert lemma i korpuset. Vi beregnet først antall forekomster av lemmaet i korpuset, delte korpuset i intervaller i forhold til frekvensen, og talte opp antall intervaller som inneholdt lemmaet. (Et forklarende eksempel kan være ordet *analyse* i et tenkt korpus på 1 million ord, der dette ordet har en frekvens på 1000. Da deles korpuset i 1000 like store tekstdele og antall tekstdele som inneholder ordet *analyse*, telles. Så telles den reduserte frekvensen per million ord i hele korpuset for å luke ut ord med mindre enn 15 i redusert frekvens per million ord.) Dette fungerte som kontroll for ord med høy redusert frekvens, men kun i en del av korpuset. Som tredje og siste steg brukte vi en stoppliste for å luke ut høyfrekvente ord som de to tidligere prosedyrene ikke fjernet.

4.4. Stopplister

Stopplister brukes for å unngå at resultatlisten ikke inneholder svært vanlige ord. Coxhead (2000) fjernet de 2000 mest frekvente ordene i The General Service List (West 1953) fra sine resultater, og Ribeck et al (2013) fjernet de

1000 mest frekvente ordene i lettlestkorpuset LäsBarT på 1 million ord. Vi antok at både størrelsen på det korpuset stopplisten ble hentet fra, og antall ord som ble tatt med på listen, ville ha betydning for resultatet, og testet det derfor ut. NoWac-korpuset, som er laget ved automatisk innsamling av tekst på .no-domenet på internett for bokmål i perioden 2009-2010 og er på 700 mill. løpeord (Guevara 2010), ble valgt som utgangspunkt. Vi brukte tre stopplister, en med de 1000 mest frekvente ordene i NoWac-korpuset, en med de 2000 vanligste, og en med de 5000 vanligste ordene i dette store internettkorpuset. Det ga tre forskjellige resultatlistene fra samme korpus.

De 500 mest frekvente ordene på alle tre resultatlistene ble gjennomgått manuelt for å undersøke om de allmenne ordene og tekniske termene var fjernet og de akademiske var med. Uaktuelle ord og ikke-ord, som navn, lånord og liknende ble talt opp. Stopplisten på 1000 ord ga 52 antatte feiltreff av de første 500 tilslagene. Det var alminnelige ord (*holdning, opplevelse*), egennavn (*Norge*), utenlandske ord og sitatord (*as, new*), tegn og tekststrukturerende elementer (*s., 2.I., ii.*) og rene tall. Stopplisten på 2000 ord ga 48 feiltreff på de 500 første tilslagene, og stopplisten med 5000 ord ga 125 feiltreff, jf. tabell 1 nedenfor.

Stoppliste	Allm.ord	Navn	Lån/sitat	Tegn	Tall	Fork.	Sum
1000 ord	28 = 5,6 %	7 = 1,4 %	8 = 1,6 %	5 = 1 %	4	0	52 = 10,4 %
2000 ord	7 = 1,4 %	18 = 3,6 %	10 = 2 %	11 = 3,4 %	1	1	48 = 9,6 %
5000 ord	7 = 1,4 %	65 = 13 %	27 = 5,4 %	22 = 4,4 %	2	2	125 = 25 %

Tabell 1: Oversikt over antall feiltreff med tre forskjellige stopplister.

Tabell 1 sier noe om hvilken stoppliste som er best egnet. Stoppliste på 1000 ord tillot relativt mange allmennord (5,6 %), mens stopplistene på 2000 og 5000 ord tok med få allmennord (1,4 %). En stoppliste på 1000 høyfrekvente allmennord er sannsynligvis for lite. Stopplisten på 5000 ord tillot en del egennavn (13 %), lånord/sitatord (5,4 %) og tegn (4,4 %), mens stoppliste på 2000 ord bare ga 3,6 % egennavn og 2,0 % lånord og 3,4 % tegn.

Vi konkluderte dermed med at en stoppliste med de 2000 mest frekvente ordene i et stort allmennspråklig korpus gir best resultat ved automatisk kartlegging av akademiske ord. Det ville være bedre å komme lavere enn

9,6 % feiltreff totalt sett, men med automatisk metode er det fortsatt vanskelig.

Vi har ikke funnet en konkret og avgrensende definisjon av 'akademisk ord' som kunne legges til grunn. Når ordet *akademisk* i seg selv er vagt, er det problematisk å gi en definisjon av hva som er et akademisk ord, og det er viktig å være klar over de metodiske problemene det innebærer. Det er derfor viktig å anvende stilistisk og pragmatisk metode i vurderingen av resultatene. De ordene som i feilanalysen anses som allmenne, har altså blitt brukt i høy grad og med stor spredning i de akademiske tekstene uten at de intuitivt ble ansett som nødvendige i beskrivelse av akademisk arbeid. Dermed utgjør de en del av den akademiske stilen. Videre mener vi at en feilforekomst eller tvilsom forekomst av "uakademiske ord" på 1,2 % heller ikke kan gjøre stor skade på en slik liste. Den automatisk genererte ordlisten må renses manuelt for feiltreff før publisering, men det er klart en fordel at det arbeidet minimeres mest mulig

4.5. Resultater for akademiske enkeltord

Analyse av det akademiske DUO-korpuset rensert med stoppliste på de 2000 mest frekvente allmennord ga etter vårt skjønn rimelig gode resultater. Som en kontroll av resultatlistene ble frekvensen i DUO sammenliknet med frekvensen i det allmennspråklige korpuset Leksikografisk bokmålskorpus (LBK) (Knudsen & Fjeld 2013). De ti mest frekvente innholdsordene på resultatlisten, ordnet etter ordklasse og med frekvenstillene i DUO og LBK er presentert i tabell 2, 3 og 4.

Det kan diskuteres i hvilken grad alle disse ordene er typisk akademiske, eller såkalt akademiske kjerneord. Intuitivt virker utvelgelsen svært god for substantiv og langt på vei for verb, mens adjektivene har flere allmenne ord som sikkert også brukes i ikke-akademiske sammenhenger. Substantivene har gjennomgående høyest frekvens, deretter verb, og på tredjeplass adjektiv. Adverbene er gjennomgående minst frekvent av de undersøkte ordene. Denne fordelingen er i samsvar med fordeling på ordklasse i norsk språk generelt. Det viser seg også at de oppslagsordene som er med i NAV, har betydelig høyere frekvens i DUO enn i LBK, som er sammensatt av forskjellige typer tekster. Sammenlikning mellom forekomst i DUO og i hele LBK-korpuset gir følgende resultater:

<i>Substantiv</i>	DUO	LBK
kapittel	2	21899
informant	3	1415
analyse	5	28778
problemstilling	7	52037
relasjon	10	1447
oppfatning	11	1082
aspekt	12	1987
funn	13	1386
kontekst	14	2070

Tabell 2: De ti mest frekvente substantiv i NAV med frekvensnummer i DUO og LBK.

<i>Verb</i>	DUO	LBK
undertrykke	8	3293
tolke	9	1302
belyse	21	2580
referere	28	2389
vektlegge	30	2922
analysere	35	2143
omhandle	37	2676
formidle	40	1626
fremstå	41	1820
tilhøre	44	823

Tabell 3: De ti mest frekvente verb i NAV med frekvensnummer i DUO og i LBK.

<i>Adjektiv</i>	DUO	LBK
teoretisk	17	1795
bevisst	18	1460
kulturell	26	1169
kvalitativ	36	3184
spesifikk	42	2558
ulik	56	2745
overordnet	83	2307
formell	95	1654
menneskelig	111	1318
kompleks	119	2488

Tabell 4: De ti mest frekvente adjektiv i NAV med frekvensnummer i DUO og i LBK.

Tallene som er understreket i tabellene, har lavere frekvens i LBK enn i NoWac, siden stopplisten er satt ved frekvens 2000 i det korpuset. Tabellene viser likevel at frekvenstallene er svært mye høyere i NAV enn i LBK, noe som viser at oppslagsordene stilistisk sett er mer typiske i akademiske tekster. Videre antar vi at det ikke bare er innholdsord (substantiv, verb og adjektiv) som er relevante for akademiske ordlister til opplæring av studenter. Resultatene viser nemlig at vår statistiske metode relativt sett fanger mange strukturord, særlig adverb, men også konjunksjoner, subjunksjoner, tekststrukturerende adverbialer og preposisjoner. Dette er ord som viser pragmatisk funksjon, som teksters logiske oppbygning eller argumentasjonsstruktur. I tillegg brukes de der det kreves i modifisert og vurdert framstilling, der man ofte må presentere sammenhengende tankerekker. Det er nettopp slik tekststruktur som er typisk for akademiske tekster, f.eks. å sette forskjellige analysedeler i sammenheng eller formulere kompliserte sammenlikninger og motsetninger eller redegjøre for forskjellen mellom kjent og ny innsikt. Spesielle adverbialer som *imidlertid*, *dessuten*, *herunder* brukes sjelden i dagligspråk, og er heller ikke nødvendig i alt fagspråk, men er pragmatisk sett nødvendige når akademiske overlegninger skal formuleres. Erfarne veiledere vet at slike ord og fraser ofte er mangelvare i mindre gode akademiske tekster.

AVL har adverbet *however* på 6. plass og SAO har preposisjonen *enligt* på 5. og konjunksjonen *samt* på 7. plass. Slike ord anvendes for å uttrykke begrunnelser, innrømmelser, forbehold eller forklarende tillegg, og det er typisk slike uttrykk man svært ofte trenger i en akademisk diskurs når man utfører det Martin (1976:92) kaller akademisk arbeid. *New Oxford Dictionary of English* definerer *however* med følgende grunnbetydning: “used to introduce a statement that contrast with or seems to contradict something that has been said previously”, og det stemmer bra i vår sammenheng.

I tabell 5 nedenfor presenteres de ti mest frekvente strukturord i DUO sammenliknet med frekvens i LBK.

<i>Adverb</i>	DUO	LBK
ibid.	24	62151
hvorvidt	33	1910
ovenfor	55	1982
underveis	86	2138
delvis	92	1325
fremfor	198	2066
således	222	2140
innad	303	5091
ifra	397	2773
innledningsvis	464	5791

Tabell 5: De ti mest frekvente adverb i NAV med frekvensnummer i DUO og i LBK.

De ti mest frekvente strukturord i AVL: 6 *however*, 32 *both*, 40 *thus*, 130 *therefore*, 151 *particularly*, 172 *indeed*, 209 *significantly*, 210 *generally*, 242 *highly*, 251 *relatively*. Og i SAO: 1 *dock*, 5 *enligt*, 15 *utifrån*, 26 *kring*, 28 *därmed*, 40 *exempelvis*, 50 *endast*, 59 *såväl*, 74 *däremot*, 77 *dels*.

Både AVL og SAO har flere og mer frekvente strukturord enn NAV. Årsaken til det kan være at strukturordene i norsk ofte omskrives til fraseologiske enheter. Vi skal derfor ta for oss en type slike flerordsenheter spesielt.

4.6. Kartlegging av flerordsenheter i akademiske tekster

Siden et vokabular ikke bare består av enkeltord, men også av fraser og andre flerordsenheter, kan videre arbeid med fokus på det fraseologiske ved akademisk språk forbedre listene. Vi har utført en avgrenset fraseologisk analyse som muligens gir bedre grunnlag for å gi svar på spørsmålet om hva et akademisk vokabular er, ut over å kartlegge enkeltord ut fra en viss frekvens i akademiske tekster. Så vidt vi har kunnet se, har arbeidet med de nordiske akademiske ordbøkene og ordlistene hittil mest fokusert på enkeltord. Resultatene i 4.5. viser at enkeltord alene ikke gir nok informasjon til å skille akademiske uttrykksmåte fra den allmennspråklige. Studier av flerordsenheter har vært sentralt i nyere leksikalsk forskning. Flerordsenheter defineres på mange måter og ut fra forskjellige perspektiv og oppgaver i språket, vi bruker her den helt nøytrale betegnelsen

flerordsenheter (engelsk Multiword expressions), som langt på vei kan erstattes med «lexical bundles» (Biber et al. 1999:990), definert som «recurrent expressions, regardless of their idiomaticity, and regardless of their structural status». Paquot har bl.a. anvendt lexical bundles (på norsk leksikalske knipper) i en studie av leksikalsk overføring fra morsmål til innlærerspråk (Paquot 2013). Vi anvender den leksikografiske termen flerordenheter, fordi leksikalske knipper antyder at det er en tettere leksikalsk sammenbinding mellom ordene enn tilfeldig samforekomst.

Vi har forsøksvis anvendt manuell analyse av bigram og trigram av tekster i DUO og sammenliknet resultatet med samme analyse av skjønnlitterære tekster for å undersøke om vi fant typiske akademiske flerordsenheter eller leksikalske knipper. Ut fra våre foreløpige resultater ser det ut til at akademisk vokabular skiller seg fra allmennvokabularet særlig på flerordsnivå. Ved hjelp av et sett metoder for automatisk seleksjon av kollokasjoner vil vi seinere undersøke mer systematisk om dette stemmer. En måte å gjøre det på, er å bruke metoden beskrevet over for bigram og trigram. En annen måte er å bruke syntaktisk avhengighets-grammatikk (semantic dependency parsing) sammen med metoden over for å lage lister basert på syntaktiske forhold. Her vil vi velge ut bare én type som en indikasjon på frasenes akademiske funksjon eller verdi.

Svært vanlige fraser i de akademiske tekstene er preposisjonsuttrykk som *med henblikk på*, *under forutsetning av* osv, ofte kalt seriepreposisjoner, bestående av preposisjon + nomen + preposisjon (her kalt PNP). Grammatisk fungerer slike ledd som regel som adverbialer, og kan erstattes av det (f.eks. *under forutsetning av* = *dersom*: *i løpet av* = *mens*). Om man vil finne slike konstruksjoner i et korpus, er det ikke tilstrekkelig med ettordsanalyser, man må søke på bi- og trigram. Vi har derfor gjort en trigramanalyse av tekster fra LBK, fordi vi der også kan sammenlikne med frekvensen i forskjellige teksttyper. LBK består av 100 millioner løpeord og har merking som gjør det enkelt å velge ut delkorpus, bl.a. etter fagområde eller teksttype. Vi etablerte derfor et sakprosa-korpus (50,3 millioner løpeord), et periodika-korpus (5,8 millioner løpeord) og et skjønnlitterært korpus (36,5 millioner løpeord) som delmengder av LBK.

I sakprosa-korpuset fant vi blant de 500 mest frekvente trigrammene 29 PNP-er, og 21 av disse forekom blant de 226 mest frekvente. Også delkorpuset periodika hadde høy frekvens av PNP, jf. tabell 6 nedenfor.

Sakprosa	Periodika	Skjønnlitteratur
2 i forhold til	6 i løpet av	7 ved siden av
6 på grunn av	10 på grunn av	46 på grunn av
11 i løpet av	12 i forhold til	59 i løpet av
14 i forbindelse med	16 i forbindelse med	257 på vei til
20 i tillegg til	22 i tillegg til	473 i nærheten av
31 i form av	60 ved universitetet i (?)	
71 i henhold til	80 til tross for	
78 i motsetning til	86 ved siden av	
84 ved siden av	155 i slutten av	
85 på bakgrunn av	161 i form av	

Tabell 6: De ti mest frekvente PNP i sakprosa, periodika og skjønnlitteratur i LBK (markert med frekvenstall) (blant de 500 mest frekvente trigram i LBK).

Til sammenlikning har vi undersøkt trigram i skjønnlitteratur. Blant de 500 mest frekvente var det bare 5 PNP i skjønnlitteratur. Disse resultatene dokumenterer at seriepreposisjoner (preposisjon+substantiv+preposisjon) er mer frekvent i sakprosa og periodika, som er de tekstene som ligger tett opp til akademisk språk, enn i skjønnlitteratur. I hvert fall er både frekvens og type svært forskjellig i de to hovedgruppene. To av de som forekom i skjønnlitteratur, kan neppe regnes til typen seriepreposisjon, da de sannsynligvis beskriver konkrete forhold: *i nærheten av* og *på vei til*.

5. Oppsummering og videre forskning

Akademiske ordlister utarbeides som hjelpemidler for studenter og ferske akademikere. Formålet er først og fremst å gi studenter hjelp ved produksjon av akademiske tekster. Basis for NAV-ordlisten er et korpus av akademiske tekster produsert ved Universitetet i Oslo på over 300 millioner løpeord. Korpuset er analysert med såkalt redusert frekvens for å finne både frekvens og spredning av potensielle akademiske ord. Resultatene ble rensset med stopplister av forskjellig størrelse for å ekskludere allmennord. Det viste seg at en stoppliste på de 2000 mest frekvente ordene i et stort internettkorpus ga resultater med færrest uønskede lemmakandidater. Vi vil seinere undersøke om en adaptert versjon av metoden til Gardner & Davis vil gi færre feiltreff enn metoden som er anvendt for SAO, slik at manuell rensing blir mindre arbeidskrevende.

En forsøksvis analyse av treordsenheter på resultatlisten viser at visse konstruksjoner med tekststrukturerende funksjon er tydelig mer frekvent i sakprosa og tidsskriftsartikler enn i romaner. En omfattende analyse av trigram i DUO-korpuset ville være interessant som sammenlikning, f.eks. bigram av adjektiv + substantiv, verb + substantiv eller leksikalske knipper av leksikalske verb og objekt, i samsvar med flere gjennomførte flerordsanalyser. Også analyser av korpus med trebankstruktur, f.eks. INESS-korpuset, som er utviklet for utforskning av syntaks og semantikk i norsk språk, der også LBK er lagt inn og analysert, vil kunne gi sikrere og tydeligere resultater om forskjeller i flerordsenheter i akademisk språk i forhold til allmennspråk.

Litteratur

- Biber, D., Johansson S., Leech, G., Conrad, S. & Finegan, E. (1999): *Longman Grammar of Spoken and Written English*. Harlow: Longman.
- Coxhead, A. (2000): A new academic word list. I: *TESOL Quarterly*, 34:2, 213-238.
- Engberg, Jan (1998): *Introduktion til fagsprogslingvistikken*. Århus: Systime.
- Gardner, D. & M. Davies (2013): A New Academic Vocabulary List. I: *Applied Linguistics* 4. <applied.oxfordjournals.org/content/early/2013/08/02/applin.amt015.full>
- Guevara, Emiliano Raul (2010): NoWaC: a large web-based corpus for Norwegian. I: *Proceedings of the NAACL HLT 2010 Sixth Web as Corpus Workshop*, Association for Computational Linguistics, 1-7.
- Hyland, K. & P. Tse (2007): Is there an "academic vocabulary"? I: *TESOL Quarterly* 41:2, 235-253.
- Knudsen, Rune Lain & Ruth Vatvedt Fjeld (2013): LBK2013: A balanced; annotated national corpus for Norwegian Bokmål. I: *Proceedings of the workshop on lexical semantic resources for NLP at NODALIDA 2013*; May 22-24 2013 Oslo. NEALT Proceedings Series 19.
- Martin, A. (1976): Teaching Academic Vocabulary to Foreign Graduate Students. *TESOL Quarterly* 10(1), 91-97

- Molde, Bertil (1976): *Fackspråk*. Skrifter utgivna av Svenska språknämnden, 57. Lund
- NAV = Norwegian Academic Vocabulary list. <www.tekstlab.uio.no:4000/>.
- Paquot, Magali (2013): Lexical bundles and L1 transfer effects. I: *International Journal of Corpus Linguistics* Vol. 18, no. 3, 391-417.
- Ralph, Bo (1981): Hur mycket fackspråk är fackspråk? I: *Svenskans beskrivning* 12. Umeå: Universitetet i Umeå.
- Ribeck, Judy, Håkan Jansson & Emma Sköldberg (2014): Från aspekt till övergripande – en ordlista över svensk akademisk vokabulär. I: Fjeld, Ruth og Marit Hovdenak (red.): *Nordiske studier i leksikografi 12. Rapport fra Konferanse om leksikografi i Norden Oslo 13.-16- august 2013*. Oslo: Novus forlag, 370-384.
- Rosén, Victoria, Koenraad De Smedt, Paul Meurer & Helge Dyvik (2012): An open infrastructure for advanced treebanking. I: Jan Hajič, Koenraad De Smedt, Marko Tadić, and António Branco (eds.): *META-RESEARCH Workshop on Advanced Treebanking at LREC2012*. Istanbul, Turkey, 22-29
- SAO = Svensk Akademisk Ordlista- (spraakbanken.gu.se/ao/om.html).
- Vinje, Finn-Erik & Helge Østbye (red.) (1981): Vanskelige ord i nyhetene (VON). En ordkunnskapsundersøkelse. (Unpubl.).
- Vinje, Finn-Erik (1982): *Journalistspråket*. Fredrikstad: Institutt for journalistikk.
- West, M. (1953): *A General Service List of English Words*. London: Longman, Green and Co.

Ruth Vatvedt Fjeld
professor
Universitetet i Oslo
Institutt for lingvistiske og nordiske studier
Boks 1102
NO-0317 Oslo
r.e.v.fjeld@iln.uio.no

Arash Saidi
vitenskapelig assistent
Universitetet i Oslo
Institutt for informatikk
arasha@student.matnat.uio.no

Å gå seg vill i en ordbok: *seg* og V+*seg* i tospråklige ordbøker

Aurelija Griškevičienė & Sturla Berg-Olsen

In this article we discuss the treatment of the reflexive pronoun *seg* and constructions containing verbs and *seg* in bilingual dictionaries with Norwegian as source language. We argue that the structure of the target language must play a key role when the lexicographer decides how to present the source language material. We propose a typology of constructions with verbs and *seg*, which serves as a basis for the treatment of these constructions in the Norwegian-Lithuanian Dictionary.

1. Innledning

Forfattere av tospråklige ordbøker må nødvendigvis ta en stor mengde prinsipielle valg når det gjelder hvordan grammatiske konstruksjoner i kilde-språket skal gjengis i målspråket. Her vil vi rette søkelyset mot hvordan konstruksjoner med verb og refleksivt pronomen i ordbøker med norsk som kilde-språk kan behandles. Vi ser først på selve pronomenet før vi analyserer verbkonstruksjoner med refleksivt pronomen i norsk og presenterer en enkel typologi for disse. Typologien kan etter vårt syn brukes som rettesnor for hvordan konstruksjonene presenteres i tospråklige ordbøker. Vi viser hvordan dette er gjort i *Norsk-litauisk ordbok* (NLO) og argumenterer for at dette systemet er konsekvent og velmotivert. Selv om vårt utgangspunkt er tospråklige ordbøker fra norsk, har diskusjonen og konklusjonene etter alt å dømme overføringsverdi til de andre skandinaviske språkene.

1.1. Prosjektet *Norsk-litauisk ordbok*

NLO er et samarbeidsprosjekt mellom leksikografer tilknyttet universitetene i Vilnius og Oslo, men prosjektledelsen holder til ved Senter for skandinaviske studier ved universitetet i Vilnius. Arbeidet med ordboka har foregått med varierende intensitet siden 2005, og hele ordboka er per i dag grov-

redigert. Det gjenstår en del kvalitetssikring før boka er publiseringsklar. Ordboka har ca. 48 000 norske oppslagsord med litauiske ekvivalenter. NLO er beregnet på brukere både med norsk og litauisk som morsmål; den er altså en bidireksjonell ordbok i Svenséns (2004:22) forstand. Ordboksmanuset foreligger i XML, og boka vil kunne publiseres både på papir og i elektronisk form.

NLO er utarbeidet med materiale fra Valerij Berkovs *Stor norsk-russisk ordbok* (SNRO) som utgangspunkt, men lemmalista fra SNRO er betraktelig redusert, og det er også gjort mange andre endringer og tilpasninger motivert av skiftet av målspråk og fordi NLO-redaksjonen på mange punkter har ønsket å ta andre valg enn Berkov. For mer informasjon om prosjektet og tilpasninger av den arvede strukturen fra SNRO se Griškevičienė & Berg-Olsen (2012) og Griškevičienė (2014).

1.2. Begreper

Vi avgrensner oss til her til behandlinga av *refleksive verb* i tospråklige ordbøker, dvs. konstruksjoner som består av et verb og et (selvstendig) refleksivt pronomen, f.eks. *kalle seg*. I norsk og andre skandinaviske språk finner man ved siden av de refleksive verbene også *s-verb*, som synkront sett består av et grunnverb og en refleksiv endelse. Diakront er *s-verbene* helt parallelle med de refleksive verbene, ettersom den refleksive endelsen er en erodert form av det refleksive pronomenet *sik* (akkusativ) eller *sér* (dativ). I moderne norsk varierer forholdet mellom refleksive verb og *s-verb* med samme grunnverb; av og til er de leksikalisert i ulike betydninger (f.eks. *spørre seg* versus *spørres*), av og til har de mer eller mindre overlappende betydning (f.eks. *skille seg* versus *skilles*).

Norsk referansegrammatikk (Faarlund, Lie & Vannebo 1997:340) bruker termen *refleksivt pronomen* bare om pronomenet *seg*. Særlig i et skandinavisk-baltisk kontrastivt perspektiv er det likevel viktig å presisere at det refleksive pronomenet i norsk realiseres likt med de personlige pronomenene når antesedenten er i første eller andre person: *hun/han/de skynder seg*, men *jeg skynder meg*, *du skynder deg* osv. Dette er annerledes f.eks. i litauisk, der det refleksive elementet forblir det samme uansett hva antesedenten er.

1.3. *Seg* i enspråklige og tospråklige norske ordbøker

Hvis vi ser bort fra de store, vitenskapelige ordboksverkene, behandler enspråklige norske ordbøker pronomenet *seg* forholdsvis overflattisk. Både *Bokmålsordboka*, *Nynorskordboka* og *Norsk ordbok med 1000 illustrasjoner* opererer med to betydningsnumre: 1) *seg* brukt i tilknytning til et verb og 2) *seg* brukt i preposisjonsuttrykk, f.eks. *gjøre mye av seg*, *noe for seg (selv)* og *være fra seg (av sinne)*. En slik behandling kan fint forsvares i ordbøker beregnet på morsmålsbrukere. De akademiske ordbøkene *Norsk riksmålsordbok* og *Norsk Ordbok* gir til gjengjeld svært detaljerte og uttømmende beskrivelser av *seg*, basert dels på syntaktiske, dels semantiske kriterier.

Det er ikke gitt at betydningsinndelinga i de enspråklige ordbøkene egner *seg* for overføring til tospråklige ordbøker. I en tospråklig ordbok bør betydningsinndelinga heller være slik at den hjelper brukerne til å finne det uttrykket i målspråket som best svarer til et visst kildepråklig uttrykk (Svensén 2004:336). Forhold i det aktuelle målspråket må altså tas med i betraktning når man velger hvordan betydningsstrukturen skal være. Dette ser vi f.eks. i Kunnskapsforlagets *Stor engelsk ordbok*, som opererer med to betydninger av *seg*: 1) (som refleksivt objekt) *herself, himself, itself, oneself, themselves*; 2) (etter preposisjon) *her, him, it, one, them*. I SNRO er ikke artikkelen **seg** delt inn i betydninger, men det er gitt en mengde eksempler som viser at norske verb med *seg* på russisk dels tilsvare verb + refleksivt pronomen, dels refleksivt verb, dels andre konstruksjoner. To tredjedeler av artikkelen i SNRO tar for *seg* uttrykk med preposisjon + *seg*, som *av seg (selv)*, *i og for seg* osv.

I NLO er det primært forhold i litauisk som dikterer strukturen i *seg*-artikkelen. Vi ser av eksemplene i ordboksartikkelen, som er gjengitt nedenfor, at betydning 1 omfatter tilfeller der litauisk har refleksive verb (med det refleksive elementet *-si-* inkorporert), mens norsk *seg* i eksemplene under betydning 2 gjerne gjengis med det litauiske refleksive pronomenet (genitiv *savęs*, dativ *sau*, akkusativ *save*, instrumentalis *savimi*).

seg I pron refl 1. (som refl. objekt i tilknytning til verb) ▶ *hun vasker seg* ji prausiasi; *han skjøt seg* jis nusišovė; *hun satte seg* ji atsisėdo; *han la seg* jis atsigulė [...] **2.** (i preposisjonsuttrykk) *savęs*; *sáu*; *savė*;

savimì ▶ *de synes ikke synd på seg selv* jiems savęs negaila; *han behóldt bóka for seg selv* jis pasiliko tą knyga sau; *han tènker bære på seg selv* jis galvoja tik apie save; *han er fornóyd med seg selv* jis patenkintas savimi [...]

1.4. Refleksive verb i SNRO

SNRO er en omfangsrik og utførlig ordbok med mange originale leksikografiske løsninger og eksempler på hvordan man kan presentere forskjellig informasjon i en ordbok, jf. omtalen i Nessel & Trosterud (2005). Når det gjelder presentasjon av refleksive verb, er det ikke gitt noen eksplisitte kriterier i innledningen av ordboka. Refleksive verb blir presentert på forskjellige måter og på ulike nivåer i ordboksartiklene, og i noen tilfeller er det bare små forskjeller mellom ulike presentasjonsmåter, f.eks. ved at det er brukt forskjellig skriftformatering. Vi kan finne de følgende ti forskjellige presentasjonsmåtene for refleksive verb i SNRO:

- 1) Når det refleksive verbet bare brukes med *seg*, kan *seg* presenteres som en del av lemmaet:¹

forlóve seg v1 **1** (*avtale ekteskap*) обруч|а́ться/-и́ться (*med – c T*) **2**
(*love for me*) сли́шком мно́го обеща́ть/=

- 2) Når *seg* er inkludert i lemmaet, kan det være skilt ut med fet skrift:

ómrømme seg v1 размышля́ть, разду́мывать

- 3) Når det refleksive verbet bare brukes med *seg*, kan *seg* være lagt inn etter lemmaet og den grammatiske koden:

skàmme II v1: ~ *seg* (*over*) стыди́ться (*P*)

- 4) Også når *seg* blir presentert etter lemmaet og den grammatiske koden, kan det være skilt ut med fet skrift:

kjåppe v1: ~ **seg** разг. торопи́ться/по-, спеши́ть/по-

¹ Artiklene fra SNRO gjengis ikke i sin helhet; informasjon som ikke er relevant for diskusjonen, er utelatt.

- 5) Når verbet brukes både med og uten det refleksive pronomenet *seg*, kan formen med *seg* være presentert som en separat betydning markert med markøren *refl.*, som i betydning 3 av **identifisere**:

identifiserle v2 **1** (*påvise identitet*) [...] **2** (*kjenne igjen*) [...] **3 refl.**:
~e *seg med* отождеств|лять/-ить себя с кем-л.

- 6) Den refleksive formen kan presenteres på betydningsnivå uten noen markør om refleksiv bruk, som i betydning 1 i av **utspille**:

utspillle v2 **1**: ~les, ~le *seg* разыгр|ываться/-аться, проис-
ходить/произойти [...] **2 idr.** [...] **3 teater** [...]

- 7) Når den refleksive formen presenteres på betydningsnivå, kan det refleksive pronomenet *seg* i tillegg framheves med fet skrift, som i betydning 3 av **forestille**:

forestille v2 **1** (*presentere*) [...] **2** (*vise*) [...] **3**: ~ **seg**
пред|ставлять/-ста́вить себе́, вообра|жать/-зítь

- 8) Når verbet brukes både med og uten det refleksive pronomenet *seg*, kan formen med *seg* være presentert som en separat del av artikkelen, som i andre del av artikkelen **oppføre**:²

oppførle v2 ① vt **1** (*bygge*) [...] **2** (*anføre*) [...] **3** (*fremføre*) [...] ②
~e *seg* вести́ себя́

- 9) Når verbet brukes både med og uten det refleksive pronomenet *seg*, kan formen med *seg* være presentert som et sublemma i slutten av artikkelen:

tènke v2 **1** [...] **6**: dét kan ~s [...]; ~ **seg** 1) (*ha planer om*) [...] 2)
(*forestille seg; anta*) [...] 3): jeg kúnne godt tènke meg å gå på kíno

- 10) Refleksive verb kan også være angitt som vanlige brukseksempler:

ánspenne v2 нап|рягать/нап|рячь; [...] ~e *seg* нап|рягаться/↑ся

² I SNRO deles artiklene i mindre enheter når verb har både transitiv og intransitiv eller refleksiv bruk og når man vil skille ut flere betydningsgrupper av preposisjoner eller adverb (f.eks. tids- og stedsbetydninger).

2. Forslag til typologi for refleksive verb og presentasjon av refleksive verb i NLO

2.1. Leksikalisert bruk

Det kan diskuteres hvor mange presentasjonsmåter av refleksive verb man trenger i en og samme ordbok, men ti måter er neppe nødvendig. En måte å sortere refleksive verb på er å analysere om forbindelsen V+*seg* utgjør en leksikalisert enhet, eller om man kan erstatte *seg* med et annet objekt uten at verbet endrer betydning. Det finnes flere typer leksikalisererte forbindelser med *seg*; vi finner det hensiktsmessig å operere med de tre følgende typene.

2.1.1. Verb leksikalisert med obligatorisk *seg*

I noen verbforbindelser med *seg* er *seg* obligatorisk, dvs. at verbet alltid brukes med *seg*, f.eks.: *forsove seg*, *late seg*, *skamme seg*. Slike verb presenteres på fire forskjellige måter i SNRO (punkt 1, 2, 3 og 4 i del 1.4.). I NLO velger vi å presentere dem med *seg* inkludert i lemmaet fordi denne måten trolig er klarest for brukeren:

ómbestemme seg v apsigalvóti

2.1.2. Verb leksikalisert med *seg*, men brukt også i andre konstruksjoner med annen betydning

Verbet kan ha én betydning når det utgjør en leksikalisert enhet med *seg*, mens det har en annen betydning når det brukes med andre objekter eller er intransitivt, f.eks: *hun reiser seg* ~ *hun reiser*, *Jens isolerer seg* ~ *Jens isolerer leiligheta*, *de drakk seg utørste* ~ *de drakk (brennevin)*, *han oppfører seg bra* ~ *han oppfører et verk*. I SNRO presenteres slike refleksive verb på fem forskjellige måter (punkt 5, 6, 7, 8 og 9 i del 1.4.). I NLO velger vi å presentere den refleksive formen i en separat del av artikkelen (del B i artikkelen **orientere**, vist nedenfor) fordi det er en bekvem måte å gjenspeile verbets semantiske struktur på når den refleksive formen har flere betydninger:

orientére v2 **A.** (vt) **1.** (anbringe) nukreĩpti, orientúoti ▶ *hùsene er orientért nord-sør* namai orientuoti iš šiaurės į pietus; *prk. en liberált orientért avís* liberalios pakraipos {orientacijos} laikraštis **2.** (opplyse) supažindinti, informúoti ▶ *han vil orientére oss om sáken* jis supažindins mus su reikalu {papasakos mums apie reikalą} **3.** idr. užsiĩmti orientávimosi spõrtu; **B.** **orientére seg** (vr) **1.** (*finne retningen*) orientúotis (*vietovėje*) **2.** (*sette inn i*) gáudytis, orientúotis ▶ *orientére seg i en ny vitenskap* gaudytis {orientuotis} naujajame moksle

2.1.3. Verb leksikalisert med *seg* og partikkel

Noen refleksive verb er leksikalisert når de brukes med en bestemt partikkel, dvs. at refleksivt verb og partikkel utgjør en fast forbindelse som har en bestemt betydning, f.eks.: *hun så seg om* ≠ *hun så seg*, *han fant seg i det* ≠ *han fant seg*, *de arbeidet seg opp* ~ **de arbeidet seg*. I SNRO blir slike forbindelser presentert enten i separate artikkelenheter (f.eks. tredje del av artikkelen **se**) eller som sublemma (f.eks. i artikkelen **finne**):

se v [...] ③ **se seg:** [...]; **se seg om** 1) осмóтрéться; 2) освóиться

finne IV v [...] □ ~ **seg i noe** терпéть/с-, по-; мирíться, примирíяться/-йт́ся с чем-л.

I NLO velger vi konsekvent å angi slike refleksive uttrykk i separate artikkelenheter merket *vr*, som i del B her:

ègle vI **A.** (vt) (su)kĩršinti, (su)érzinti, (su)kùrstyti, (iš)provokúoti ▶ *ègle (noen) opp* sukĩršinti, išprovokuoti (*kq*); **B.** **ègle seg** (vr) ◆ **ègle seg innpå** *noen* kabinétis, priekabiáuti (*prie ko*)

2.2. Ikke-leksikalisert bruk

Når det gjelder ikke-leksikalisererte tilfeller av V+*seg*, kan vi skille ut to hovedtyper, som vi diskuterer hver for seg nedenfor.

2.2.1. Ikke-leksikalisert bruk med verb som brukes med samme betydning også i andre konstruksjoner

Den første typen inneholder verb som også brukes intransitivt eller med andre objekter, og da med samme betydning som i konstruksjonen med *seg*. Noen eksempler på dette er *hun vasker seg* ~ *hun vasker (mannen sin)*, *de la seg* ~ *de la ungene* og *de sminket seg* ~ *de sminket hverandre*. Igjen ser vi at SNRO bruker flere ulike presentasjonsmåter, for **kjemme** er f.eks. den refleksive bruken skilt ut i en egen artikkeldel:

kjømme II v1 ① vt **1** [...] **2** [...] ② ~ **seg** расчѣс|ываться/-аться; причѣсываться/↑

I NLO gis informasjon om den refleksive bruken i disse tilfellene som vanlige brukseksempler:

kjømme II v1 // *kjømte, kjemt* **1.** (gre) šukúoti ▶ *kjømme håret* šukuoti plaukus; *kjømme seg* šukuotis [...]

2.2.2 Ikke-leksikalisert brukt der *seg* uttrykker benefaktiv eller malefaktiv

En annen type ikke-leksikalisert konstruksjon av typen V+*seg* har vi når *seg* uttrykker en benefaktiv (en som verbhandlingen er til fordel for) eller malefaktiv (en som verbhandlingen er til ulempe for). I SNRO kan man f.eks. finne eksempler på dette angitt i en egen betydningsbolk:

ópptjene v2 **1** (*skaffe seg rett til*) [...] **2:** ~ **seg** en stor fòrmue нажítъ большóе состоя́ние

I den grad bruk av *seg* for benefaktiver og malefaktiver gis i NLO, står også disse som alminnelige brukseksempler:

ópptjene v2 užsidirbti, užsitarnáuti ▶ *ópptjene álderstillegg* užsidirbti {užsitarnáuti} priedą už darbo stažą; *ópptjene fòrskudd* atidirbti už avansą; *ópptjene seg en stor fòrmue* susikrauti didelį turtą

3. Konklusjon

Vi har sett at behandlinga både av det refleksive pronomenet *seg* og verbforbindelser med *seg* er grunnleggende forskjellig i enspråklige ordbøker beregnet på morsmålstalere og tospråklige ordbøker som skal kunne brukes av talere av begge språk. Dette er i tråd med prinsippet om at strukturen i målspråket må tas med i betraktning når artiklene i tospråklige ordbøker utformes. Ikke desto mindre er det god grunn til å nærme seg konstruksjoner av typen *V+seg* på en systematisk måte og ikke overdrive antallet presentasjonsmåter. Vi har presentert en enkel typologi for *V+seg* i norsk som vi mener utgjør et godt utgangspunkt for behandlinga av disse konstruksjonene i tospråklige ordbøker. Typologien er basert dels på hvorvidt *V+seg* utgjør en leksikalisert enhet eller ikke, og dels på hvorvidt verbet har samme betydning brukt med *seg* som i andre konstruksjoner. På grunnlag av denne typologien opererer vi i NLO med tre ulike presentasjonsmåter, noe som etter vårt syn både er teoretisk velfundert og brukervennlig.

Litteratur

Ordbøker

- Bokmålsordboka* (2005): Boye Wangensteen (red.). 3. utgave. Oslo: Kunnskapsforlaget. <www.nob-ordbok.uio.no/perl/ordbok.cgi?OPP=&bokmaal=+&ord-bok=bokmaal> (september 2015).
- Norsk Ordbok. Ordbok over det norske folkemålet og det nynorske skriftmålet* (1966–2016): Oslo: Det Norske Samlaget. <no2014.uio.no> (januar 2016).
- Norsk Ordbok med 1000 illustrasjoner* (2005): Tor Guttu (red.) 2. utgave. Oslo: Kunnskapsforlaget.
- Norsk riksmålsordbok* (1937–1995): Trygve Knudsen & Alf Sommerfelt (red.). Oslo: Kunnskapsforlaget.

- Nynorskordboka* (2006): Marit Hovdenak et al. (red.). 4. utgåva. Oslo: Det Norske Samlaget. <www.nob-ordbok.uio.no/perl/ordbok.cgi?OPP=&nynorsk=+&ordbok=nynorsk> (september 2015).
- SNRO (2003) = Valerij P. Berkov: *Stor norsk-russisk ordbok*. Oslo: Kunnskapsforlaget.
- Stor engelsk ordbok*. Oslo: Kunnskapsforlaget. <www.ordnett.no/butikk/engelskeordboker.storengelskordbok> (september 2015).

Annen litteratur

- Faarlund, Jan Terje, Svein Lie & Kjell Ivar Vannebo (1997): *Norsk referansegrammatikk*. Oslo: Universitetsforlaget.
- Griškevičienė, Aurelija (2014): Jolly and less jolly challenges in working on the Norwegian-Lithuanian Dictionary. I: Ērika Sausverde & Ieva Steponavičiūtė (eds.): *Fun and Puzzles in Modern Scandinavian Studies = Scandinavistica Vilnensis* 9. Vilnius: Vilnius University, 27–40.
- Griškevičienė, Aurelija & Sturla Berg-Olsen (2012): A golden mean? Compromises between quantity of information and user-friendliness in the bi-directional Norwegian-Lithuanian Dictionary. I: Ruth Vatvedt Fjeld & Julie Matilde Torjusen (red.): *Proceedings of the 15th EURALEX International Congress*. Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo, 648–653.
- Neset, Tore & Trond Trosterud (2005): Ny norsk-russisk ordbok: Ei leksikografisk storhending. I: *Lexiconordica* 12: 273–284.
- Svensén, Bo (2004): *Handbok i lexikografi*. Stockholm: Norstedts Akademiska Förlag.

Aurelija Griškevičienė
dosent, hum. m. dr.
Senter for skandinaviske studier
Vilnius universitet
Universiteto 3
LT-09314 Vilnius
aurelija.griskeviciene@flf.vu.lt

Sturla Berg-Olsen
seniorrådgiver, dr.art.
Språkrådet
Postboks 8107 Dep
NO-0032 Oslo
sturla.berg.olsen@sprakradet.no

Kodifiering av multietniskt ungdomsspråk

Jon Helgason & Emma Sköldberg

This article covers some of the findings of a pilot study of what we perceive to be a tendency towards a codification of the contemporary urban vernacular in written Swedish. In Sweden, the urban vernacular has traditionally been described as a youth language, and, as such, studied as a part of an (ephemeral) youth culture. However, both internationally and in Sweden there are claims to treat the urban vernacular as an urban dialect or ethnolect. This article contains a brief presentation of some the materials of this pilot study, as well as an account of three lexical items and how they are described in (mostly user-generated) lexicographical resources.

1. Inledning

En tendens som uppmärksammas i svensk och internationell forskning är att s.k. multietniska ungdomsspråk (MUS) alltmer antar formen av ”urbana dialekter” eller etnolekter i storstädernas mer invandrartäta områden (Bodén 2010; Wiese 2009). I vårt bidrag lyfter vi fram ansatser till språklig kodifiering i skrift av det aktuella ordförrådet på svensk mark, något som kan sägas utgöra en form av validering av ovan nämnda tendens. Syftet med bidraget är tudelat: för det första presenteras ett material som kan ligga till grund för studier av den pågående kodifieringen av MUS. För det andra redovisas en i första hand kvalitativt inriktad och inventerande förstudie av bruket av några MUS-ord.

2. Bakgrund och utgångspunkter

De senaste decennierna har åtskillig forskning ägnats åt multietniskt ungdomsspråk, såväl i Sverige som internationellt (se Rampton 1995, 2011; Nortier 2001; Androutsopoulos & Georgakopoulou 2003; Jaspers 2005; Madsen 2008; Källström & Lindberg 2011; Bijvoet & Fraurud 2013 m.fl.). I

detta sammanhang bör också Ulla-Britt Kotsinas lyftas fram som en pionjär genom sina sociolingvistiska studier av *Rinkebysvenska* alltsedan 1980-talet.

Multietniskt ungdomsspråk är den påbjudna språksociologiska benämningen av detta fenomen, som också har många andra namn: *förortsslang*, *miljonsvenska*, *gringosvenska*, *blattesvenska* (jfr Källström 2011). I Danmark har Pia Quist (2000) föreslagit termen *multietnolekt* för att signalera att det hänger samman med en erkänd språklig varietet som beskrivits med andra lekt-begrepp (dialekt, sociolekt).

I Sverige har studier av MUS i huvudsak följt två huvudfåror. Inom en fåra har man fokuserat på MUS som språklig avvikelse i relation till standardspråket. Man har studerat grammatiska egenheter, särskilt s.k. grammatisk reduktion avseende syntax och morfologi (Ganuza 2010) och uttal (Bodén 2007). I förhållande till standardsvenskan har MUS en omarkerad och förenklad syntax och morfologi. Typiska grammatiska drag är strykningen av betydelsestomt ”det” som objekt eller subjekt, samordning av tidsböjda verb utan ett ”och”, konstruktion med direkt objekt istället för med verb och preposition och avvikande bruk av bestämd och obestämd artikel (Kotsinas 2005:243). Man har även studerat hur vissa vanligt förekommande ord- eller ordled tjänar som förstärkningsord eller som emfatiska uttryck, särskilt med funktion som diskurspartiklar (se t.ex. Svensson 2009). Inom den andra fåran har man studerat de språkattityder och identitetskonstruktioner som den språkliga variationen är förbunden med, samt samtida sociala, ideologiska och politiska föreställningar förknippade med MUS. Ytterst har denna fåra ifrågasatt föreställningen om MUS som språklig avvikelse och istället betonat hur detta perspektiv hänger samman med frågeställningar om normativitet och makt (t.ex. Stroud 2004, 2013), något som förklarar MUS politiska och sociala laddning i offentligheten (Källström 2011; Milani 2013).

Vår undersökning knyter snarare an till det som stundom kallas ”den tredje vågen” inom sociolingvistiska variationsstudier (Eckert 2012). Vi betraktar inte språket enbart som en reflektion av sociala identiteter och kategorier utan vi vill även betona språkanvändarnas agens. I nyare forskning granskas och analyseras språket som social praktik och som menings- och identitetsmarkör (Jørgensen & Quist 2008; Opsahl 2015). I detta perspektiv utgör språket ett symbolverktyg som kan skapa och upprätthålla sociala kategorier och gränser.

3. Material och metod

En komplicerande metodologisk omständighet vid flertalet studier av MUS är att det i huvudsak rör sig om muntligt språkbruk. Det innebär bland annat att orden, när de förekommer i skriftspråk, inte har en etablerad ortografi. Vi har valt att studera hur ord som traditionellt brukas föras till MUS används eller beskrivs i fem typer av material. Sammanfattningsvis är de undersökta materialen av olika och kompletterande slag:

1) Svenska raptexter (från perioden 1994–2011). Ett urval texter har med hjälp av konkordansprogrammet AntConc satts samman i en korpus (på i nuläget drygt 100 000 löpande ord).¹ De transkriberade texterna är hämtade från olika nätresurser där låttexter troligen oftast har nedtecknats av andra än upphovspersonerna.

2) Svensk skönlitteratur (i nuläget tolv verk, utgivna mellan åren 2000–2015). Av praktiska och upphovsrättsmässiga skäl har detta delmaterial excerperats manuellt. Vår inventering av skönlitterära förekomster gör inga anspråk på att vara uttömmande, men av den framgår att MUS förekommer i så gott som alla genrer (i vid mening). Det är representerat i kriminallitteratur, såsom Jens Lapidus *Snabba Cash* (2006) och Rouzbeh Masarrat Agahs *FTL* (2013) och i ungdomslitteratur genom Douglas Foleys bokserie *Shoo bre* (2003), *Shoo len* (2009), *Shoo mannen* (2012). Det uppträder också i barnböcker genom Dogge Doggelito & Yoria Riveras bokserie om Izzy (2011, 2011, 2012). Inslag av MUS återfinns också i s.k. kvalitetslitteratur genom Alejandro Leiva Wegners *Till vår ära* (2001), och Hassan Loo Sattarvandis *Still* (2008), båda utgivna på Albert Bonniers förlag. Avslutningsvis finns det även i självbiografiska framställningar, t.ex. Sebastian Staksets biografi *Sebbe Staxx* (2015) och i litterära omgestaltningar som Dimitrios Iordanoglous *Illiaden – en cover* från 2010.

3) Miniordlistor. Vissa av ovan nämnda romaner innehåller korta ordlistor med förklaringar av ett flertal MUS-ord som förekommer i verken. Listornas funktion är inte självklar, men de kan givetvis fungera som stöd vid läsning av romanerna. Vissa författare, som Dogge Doggelito & Yoria

¹ Urvalet gjordes primärt utifrån ambitionen att täcka den angivna perioden, 1994–2011.

Rivera, tycks också vara måna om att lyfta fram eller höja statusen hos denna typ av vokabulär.

4) Tryckta ordböcker. I detta delmaterial ingår dels allmänspråkliga ordböcker som SAOL, *Natur och Kulturs Stora Svenska ordbok* (2006), *Svensk ordbok utgiven av Svenska Akademien* (2009) och *Bonniers svenska ordbok* (10:e uppl., 2010), dels slangordböcker som *Norstedts svenska slangordbok* (1998) och *Förortsslang* (2004).

5) Användargenererade lexikografiska resurser. Dit räknar vi bl.a. Wiktionary, Slangopedia.se, Tyda.se och Folkmun.se.

Materialet hade kunnat utökas, till exempel genom att även inkludera texter från sociala medier, exempelvis blogginlägg. Det får emellertid vänta till en större, mer omfattande studie.

4. Tre exempel

Det material som undersöks är mycket omfattande och inom ramen för denna studie finns endast utrymme för några nedslag. Vi utgår från tre exempel: substantiven *guss* och *shonu* med alternativstavningar och verbet *gitta/git*. Enligt *Norstedts svenska slangordbok* (1998) härstammar samtliga ord från turkiskan.

4.1. Substantivet *guss/guzz*

I vårt material är första belägget på substantivet *guss* i en låttitel på Latin Kings debutalbum *Välkommen till förorten* (1994). I en recension av detta album förefinns även ett av de första skriftliga beläggen för ordet *guss* (*Expressen* 13 maj 1994). När de olika formerna i raptexterna har lemmatiserats finns där totalt 55 *guss* och 48 *guzz*.

I den skönlitterära korpusen har emellertid formen *guzz* tagit överhanden, även om varianten *guss* också förekommer. I en av miniordlistorna förekommer ännu en form, nämligen *güzz* (med betydelseangivelsen 'tjej'). En intressant iakttagelse är att detta ord inte är upptaget i *FTL*:s relativt utförliga miniordlista, trots att det är vanligt förekommande i romanen. Men listan saknar också *aina* ('polis'), *keff* ('dålig') och *jalla* ('skynda, kom igen'),

som också används flitigt i verket. En tänkbar förklaring är att författaren betraktar dessa ord som etablerade.

I tryckta ordböcker behandlas ordet först i *Norstedts slangordbok* (1998), som redovisar fyra olika varianter: *gus*, *guss*, *guz* och *guzz*. Två huvudbetydelser anges: 'flicka' och 'flickvän'. Nästan 10 år senare, år 2006, kommer *guss* in i SAOL och träder därmed in i det lexikografiska finrummet. I ordlistan anges en form, svenskklingande *guss*, och att ordet tillhör ett vardagligare språkbruk. Ordet är försett med böjningsangivelserna *-en*, *-er* och betydelsen uppges som 'flicka'. Andra ordböcker, t.ex. *Bonniers svenska ordbok* och *Natur och Kulturs Stora svenska ordbok* följer sedan efter med liknande artiklar. Det faktum att två MUS-ord (*guss* och ovan nämnda *keff*) fanns med bland de ca 10 000 nyorden i den trettonde upplagan av SAOL väckte livlig debatt. Se t.ex. Nikula (2010) som tar upp ett blogginlägg med rubriken "SAOL våldtagen".

Avslutningsvis behandlas *guss/guzz* i en rad användargenererade lexikon, som Tyda.se, Luxikon, Folkmun.se samt även i nätresurser som Öresunds översättningsbyrås ordlista. Behandlingen av ordet i svenska Wiktionary illustrerar hur bidragsgivarna på den här typen av sajter interagerar vid beskrivningen av MUS-ord. I diskussionstråden som är knuten till artikeln, tas nämligen angivelsen av just detta ords stavning, böjning och stilnivå upp. Denna form av interaktion vill vi framhålla som en kännetecknande del av kodifieringen.

4.2. Substantivet *shonu* med alternativstavningar

Bruket av *shonu* är mer heterogent och därmed svårare att ringa in. I vår rapkorpus förekommer ordet sammanlagt 94 gånger. Efter en lemmatisering framträder inte mindre än sju stavningsvarianter: *chuno* (1 belägg), *shone* (14), *shonne* (33), *shouno* (2), *shune* (11), *shunno* (1) och *shuno* (32).

Också i skönlitteraturen är stavningsvariationen påfallande även om den inte är lika stor. I något verk får man också hjälp med vad ordet betyder. I barnboken *Izzy & gänget* (2011:30; kursiv i original) står det: "Torget är hjärtat av betongen (...). Där kan man spana på alla coola *shunos* och *güzzar* som glider i de fetaste bilarna man kan tänka sig." Originallets kursivering indikerar att orden ingår i verkets miniordlista. Där står det: *shuno* ('snubbe'). I en annan ordlista, den i *FTL*, återfinns uppslagsordet *shonno*,

med betydelsen 'han', 'hon'. Ordet tycks alltså också fungera som ett pronomen som kan referera till både män och kvinnor.

Shonu med alternativstavningar redovisas inte i de allmänspråkliga ordböckerna i vårt material, men däremot i *Norstedts slangordbok*. Ordboksartikeln påvisar en omfattande ortografisk variation genom att inte mindre än fem varianter redovisas: *shono*, *shonu* (äv. *chono*, *chuno*, *shuno*). En mängd betydelser anges. Betydelsen uppges vara såväl 'flicka, flickvän' som 'pojke, pojkvän'. Ordet anges också fungera som pronomen (med betydelsen 'han/hon'). Som sista betydelse anges 'narkoman'.

Det aktuella ordet behandlas i endast en av de användargenererade ordböckerna i materialet, Slangopedia.se. Signaturen "Anonym" lade i februari 2011 in uppslagsordet *shuno* följt av kommentaren "ett annat ord för kille, snubbe". På sajten finns ett lätt humoristiskt språkprov i form av en fiktiv dialog som anges utspela sig i förorten: – *Ey vem fan e den där shunon? – Ingen aning mannen.* Detta språkprov följs sedan av en normalspråklig återgivning av samma dialog: – *Karl-Göran, vem är den där killen egentligen? – Har inte den blekaste aning.* Som synes används ett mer vårdat språk i det senare fallet.

4.3. Verbet *gitta/git*

Det tredje och sista exemplet gäller verbet *gitta*, som liksom *shuno* uppträder i en rad former i raptexterna. Efter lemmatisering rör det sig om totalt 33 belägg i vår rapkorpus (*gitta* 14, *gitt* 11, *gittar* 5, *gittat* 1 samt *gittish* 2).

Olika former av ordet förekommer också i de skönlitterära verken, däribland Foleys *Shoo bre*:

- (1) Vi kan gitta till centrum... (s. 42)
- (2) Gitt len! Fattar inte hur du tänker ibland. (s. 19)
- (3) Jalla, vi gitt till centrum, bre. (s. 21)
- (4) Var är dina polare? (...) Dom gittish hemåt (s. 29)

Ordet torde betyda 'sticka, dra' i samtliga fall. Författaren tycks dock ta för givet att läsaren antingen redan känner till ordet eller kan räkna ut betydelsen utifrån kontexten.

Endast en av de tryckta ordböckerna, *Förortssvenska*, behandlar ordet, men av någon anledning behandlas det som tre uppslagsord, *git* ('gå din

väg!, stick!'), *gitta* ('gå, sticka, ge sig iväg') och *gittish* ('springa, sticka'). I det tredje fallet har grundformen försetts med den MUS-typiska produktiva men semantiskt bleka slutleden *-ish*. Ordet är däremot oftare behandlat i nät-resurser av olika slag. Också i detta fall utgör brukarnas kommentarer intressant läsning för den som vill få djupare kunskaper om ordens form, betydelse och grad av etablering. Artikeln *gitta* i Slangopedia.se kan utgöra exempel på det. Enligt signaturen "Bibido Zratman", som skapade artikeln 2008, betyder det "Att gå iväg, sticka, dra sin kos, tagga, försvinna eller pysa". Signaturen "pewi" ifrågasätter denna betydelsebeskrivning och säger att betydelsen är 'orka'. Signaturen "Sofia" responderar: "Öhh gitta (git) betyder dra/gå på turkiska. Det betyder fan inte 'orka' lol". Signaturen "Rebecca" stöder däremot "pewi". Rebecca anser också att ordet betyder 'orka' och hänvisar till SAOL. Signaturen "miwi" försöker bringa klarhet i diskussionen och konstaterar: "det är två helt olika ord. så ni båda har rätt." "Don Pedro" avslutar diskussionen genom att konstatera: "Slang-gitta uttalas med hårt g och SAOL-gitta med mjukt g."

5. Slutord

Den här artikeln behandlar den pågående kodifieringen av multietniskt ungdomsspråk (MUS) i skrift. För det första presenteras ett material, som kan tjäna som utgångspunkt för studier av denna process. För det andra redovisas delar av en pilotstudie, omfattande tre ord som brukar föras till MUS.

I samband med att MUS förekommer transkriberat eller i skriftspråk har en form av normalisering eller standardisering ägt rum. I de lexikografiska resurserna anges ofta alternativa former och hänvisningarna är många. Den som söker på MUS-ord har ofta bara hört dem sägas och uttalet kan, som för dialekter i allmänhet, ligga rätt långt från den normaliserade uppslagsformen (i den mån en normaliserad variant finns) (Hallén & Thelin 2006). I samband med denna normalisering (kodifiering) tycker vi oss skönja två ansatser, dels en anpassning av orden till svensk stavning och morfologi, dels den att behålla eller understryka ordets främmande karaktär.

De tre exempelorden, *guss*, *shonu* och *gitta*, behandlas på olika sätt i de ordböcker och lexikografiska resurser som vi har studerat. Medan formen *guss* har letat sig in i så gott som samtliga verk, tycks såväl *shonu* som *gitta*

vara mindre kända och spridda. Dessa ord behandlas dock i flera användargenererade ordböcker – resurser som spelar en viktig roll vid studier av vardagligare ord och uttryck.

En komplicerande metodologisk omständighet i vår studie är att de studerade orden, när de förekommer i skriftspråk, inte har en etablerad ortografi. En fråga som vår sonderande undersökning väcker är om detta ordförråd överhuvud vill bli fångat och kodifierat (jfr Opsahl 2015). Vill de personer som mer frekvent använder dessa ord och uttryck att orden inkluderas i svenskt standardspråk? Eller vill de att denna varietet ska bibehålla sin funktion som identitetsmarkör – en form av utanförskap? Ett exempel som illustrerar denna frågeställning är substantivet *guss*. När ordet först uppträdde i skrift stavades det, enligt våra material, främst *guss*. Denna form blev också upptagen i flera svenska ordböcker inklusive SAOL. Som bidragande skäl för upptagandet angavs bland annat att ordet lätt kunde passas in i standardspråkets uttals- och böjningsregler. Därefter kom formen *guzz* starkt. Denna stavningsvariant är för tillfället vanligast, åtminstone i våra material. På senare tid uppträder en ny form, *güzz*, som tycks ta fasta på ett y-uttal. Det ska bli spännande att se om denna variant etableras ordentligt – och hur länge den i sådana fall består.

Litteratur

- Androutsopoulos, Jannis, & Alexandra Georgakopoulou (red.) (2003): *Discourse Constructions of Youth Identities*. Amsterdam: John Benjamins.
- Bijvoet, Ellen & Kari Fraurud (2013): "Rinkebysvenska" och andra konstruktioner av språklig variation i dagens flerspråkiga Sverige. I: Kenneth Hyltenstam & Inger Lindberg (red.): *Svenska som andraspråk – i forskning, undervisning och samhälle*. (2 uppl.). Lund: Studentlitteratur, 369-396.
- Bodén, Petra (2007): "Rosengårdssvensk" fonetik och fonologi. I: Lena Ekberg (red.): *Språket hos ungdomar i en flerspråkig miljö i Malmö*. (Nordlund 27. Småskrifter från Nordiska språk.). Lund: Nordiska språk, Språk- och litteraturcentrum, 1-47.

- Bodén, Petra (2010): Pronunciation in Swedish multiethnolect. I: Pia Quist & Bente Ailin Svendsen (red.): *Multilingual Urban Scandinavia. New Linguistic Practices*. Bristol: Multilingual Matters, 65-78.
- Eckert, Penelope (2012): Three Waves of Variation Study: The Emergence of Meaning in the Study of Sociolinguistic Variation. I: *Annual Review of Anthropology* 41, 87-100.
- Ganuza, Natalia (2010): Subject-Verb Order Variation in the Swedish of Young People in Multilingual Urban Areas. I: Pia Quist & Bente Ailin Svendsen (red.): *Multilingual Urban Scandinavia. New Linguistic Practices*. Bristol: Multilingual Matters, 31-48.
- Hallén, Karin & Eva Thelin (2006): Ett svenskt dialektlexikon i startgrupparna. I: Henrik Lorentzen & Lars Trap Jensen (red.): *Nordiske Studier i Leksikografi* 8. Köpenhamn, 163-170.
- Jaspers, Jürgen (2005): Linguistic sabotage in a context of monolingualism and standardization. I: *Language and Communication* 25/3, 279-298.
- Jørgensen, J. Normann & Pia Quist (2008): *Unges sprog*. Köpenhamn: Hans Reitzels Forlag.
- Kotsinas, Ulla-Britt (2005): *Invandrarsvenska*. Uppsala: Hallgren och Fallgren.
- Källström, Roger (2011): Multiethnic youth language in reviews of the novel *Ett öga rött*. I: Roger Källström & Inger Lindberg (red.): *Young Urban Swedish. Variations and change in multilingual settings*. Göteborg: Institutionen för svenska språket, Göteborgs universitet, 125-147.
- Källström, Roger & Inger Lindberg (red.) (2011): *Young Urban Swedish. Variations and change in multilingual settings*. Göteborg: Institutionen för svenska språket, Göteborgs universitet.
- Madsen, Lian Malai (2008): *Fighters and Outsiders: Linguistic Practices, Social Identities, and Social Relationships among Urban Youth in a Martial Arts Club*. Köpenhamns universitet: Institut for Nordiske Studier og Sprogvidenskab.
- Milani, Tommaso M. (2013): Språkideologiska debatter i Sverige. I: Kenneth Hyltenstam & Inger Lindberg (red.): *Svenska som andraspråk – i forskning, undervisning och samhälle*. (2 uppl.). Lund: Studentlitteratur, 343-367.

- Nikula, Kristina (2010): I språkets utförsbacke – om icke-lingvistisk språkkritik. I: Kristinn Jóhannesson m.fl. (red.): *Bo65. Festskrift till Bo Ralph*. (Meijerbergs arkiv för svensk ordforskning 39). Göteborg, 343-352.
- Nortier, Jacomine (2001): *Murks en Straattaal*. Prometheus: Amsterdam.
- Opsahl, Toril (2015): Kan ord i bruk bli ord i bok? Urbane ungdomsvarianteter i framtida ordboksressurser I: *LexicoNordica* 22, 131-149.
- Quist, Pia (2000): Ny københavnsk 'multietnolekt'. Om sprogbrug blandt unge i sprogligt og kulturelt heterogene miljøer. I: *Danske Talesprog* 1, 143-212.
- Quist, Pia (2013): Skandinavisk i kontakt med indvandrersprog. I: *Språk i Norden 2012: Tema: Morsmål, andrespråk, nabospråk og fremmedspråk*. Utgitt av Nettverket for språknemndene i Norden, 26-37.
- Rampton, Ben (1995): *Crossing: language and ethnicity among adolescents*. London: Longman.
- Rampton, Ben (2011): From 'Multi-ethnic urban heteroglossia' to 'Contemporary urban vernaculars'. I: *Language & Communication* 31, 276-294.
- Stroud, Christopher (2004): Rinkeby Swedish and Semilingualism in Language Ideological Debates. A Bourdieuan Perspective, *Journal of Sociolinguistics* 8, 196-214.
- Stroud, Christopher (2013): Halvspråkighet och rinkebysvenska som språkideologiska begrepp. I: Kenneth Hyltenstam & Inger Lindberg (red.): *Svenska som andraspråk – i forskning, undervisning och samhälle*. (2 uppl.). Lund: Studentlitteratur, 313-342.
- Svensson, Gudrun (2009): *Diskurspartiklar hos ungdomar i mångspråkiga miljöer i Malmö*. (Lundastudier i nordisk språkvetenskap A 67). Lunds universitet: Språk- och litteraturcentrum.
- Wiese, Heike (2009): Grammatical innovation in multiethnic Europe: New linguistic practices among adolescents. I: *Lingua* 119, 782-806.
- Wölck, Wolfgang (2002): Ethnolects – between bilingualism and urban dialect. I: Li Wei, Jean-Marc Dewaele & Alex Housen (red.): *Opportunities and Challenges of Bilingualism*. Berlin/New York: Mouton de Gruyter, 157-170.

Jon Helgason
forskare, fil. dr
Språk- och litteraturcentrum
Lunds universitet
Box 201
SE-221 00 Lund
jon.helgason@litt.lu.se

Emma Sköldberg
universitetslektor, docent
Institutionen för svenska språket
Göteborgs universitet
Box 200
SE-405 30 Göteborg
emma.skoeldberg@svenska.gu.se

SAOLhist.se – för allmänt och vetenskapligt bruk

Louise Holmer, Sven-Göran Malmgren & Monica von Martens

The Swedish Academy Glossary (Svenska Akademiens ordlista, SAOL), a normative glossary with particular focus on spelling, inflexion, and pronunciation, has appeared since 1874. In 2006, the 13th edition appeared. The second, third, fourth and fifth editions were more or less reprints of the first edition. Thus, the 2006 edition may be regarded as the ninth (autonomous) SAOL edition. The most important parts of these nine editions, especially the complete sets of lemmas, have been transformed into a database. The database may be thought of as a matrix with more than 200 000 rows and eleven columns (the nine SAOL editions plus the Dalin dictionary, plus a column to the left with all lemmas that have occurred in at least one SAOL edition). The interface is very flexible, allowing the inclusion of other dictionaries. From each square of the matrix there is a link to a facsimile version of the full corresponding dictionary article. Several advanced searches are possible, e.g.: which words were added and omitted between any two consecutive editions, or: has the frequency of a particular suffix increased or decreased over time?

1. Inledning

Sedan 2013 finns databasen SAOLhist fritt tillgänglig på nätet på adressen saolhist.se. Databasen innehåller uppslagsorden (med ordklassangivelser) i nio av de 13 upplagor av *Svenska Akademiens ordlista* (SAOL) som hade kommit ut 2012. Det tidiga arbetet med bland annat digitaliseringen av de äldre upplagorna beskrivs närmare i Holmer (2012). Upplagorna 2–5 var i stort sett omtryck av den första upplagan och uteslöts. De upplagor som ingår i SAOLhist är sålunda SAOL 1 (1874), SAOL 6 (1889), SAOL 7 (1900), SAOL 8 (1923), SAOL 9 (1950), SAOL 10 (1973), SAOL 11 (1986), SAOL 12 (1998) och SAOL 13 (2006) (jfr t.ex. Malmgren 2014). Sedan april 2015 föreligger även SAOL 14, såväl i pappers- som databasform. Den ingår i en intern version av SAOLhist men ännu inte i den allmänt tillgängliga versionen.

Databasens struktur och det flexibla gränssnittet gör att man lätt kan komplettera den med data från andra ordböcker. Sålunda ingår även lemmauppsättningen i Dalins klassiska ordbok (1850–53) i SAOLhist. Ett samarbete har också inletts med Dansk sprogævn, vilket resulterat i att alla upplagor av den danska *Retskrivningsordbogen* (sedan 1955) nu föreligger i en ”systerdatabas” med ett modifierat gränssnitt (rohist.dsn.dk).

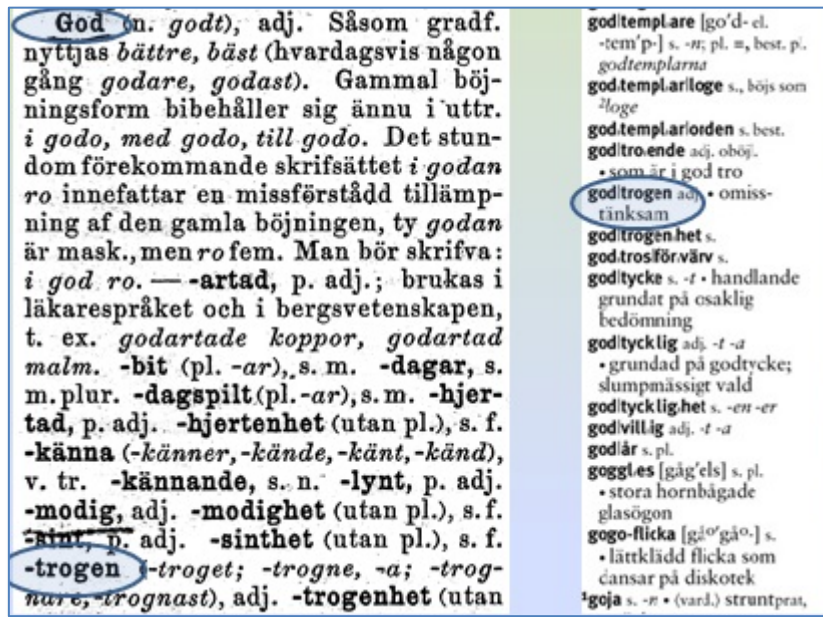
I det följande ger vi först en kort redogörelse för arbetet med att etablera SAOLhist, med fokus på de mest intressanta lexikografiska problemen. Sedan följer en diskussion av de datalogiska aspekterna på databasen. Slutligen ges några exempel på hur man kan använda SAOLhist – både som forskare och som allmänt språkintresserad.

2. Etableringen av SAOLhist: några lexikografiska problem

SAOLhist-projektet fokuserade från början på lemmauppsättningen, jämte informationen om ordklassstillhörighet. Den övriga informationen, t.ex. uttals- och böjningsinformationen och den i tidigare upplagor ytterst rudimentära betydelseinformationen, negligerades. Tre av de nio SAOL-upplagorna förelåg redan i digital form, nämligen upplagorna 11–13. De övriga skannades in och OCR-tolkades med fokus på att möjliggöra maskinell extrahering av feta teckensträngar (=uppslagsord) och (nästan) omedelbart följande ordklassangivelser, varefter normalisering och en första korrekturläsning genomfördes. Rättning och korrekturläsning utfördes till stor del av tredje- eller fjärdeterminsstuderande i nordiska språk. I ett senare skede gjordes mer sofistikerade korrekturläsningar, varvid bl.a. de garanterat korrekta uppslagsformerna i upplagorna 11–13 utnyttjades som jämförelsematerial.

De elva första SAOL-upplagorna tillämpade en nischalfabetisk modell med huvudlemman och sublemman (se Figur 1). Det innebar att de flesta uppslagsord inte återgavs explicit utan med hjälp av en divis som representerade ett grundord eller en förled i en sammansättning. Formalismen för att särskilja sammansättningar och avledningar utvecklades gradvis från en utgåva till nästa, vilket medförde att metoden för digitalisering och extrahering måste anpassas för varje separat utgåva. Efterlederna i de tidiga upplagorna

expanderades algoritmiskt till fullständiga uppslagsord för att göras jämförbara med uppslagsorden i de senare upplagorna, se t.ex. **godtrogen** i Figur 1.



Figur 1: Artiklarnas uppställning i SAOL 1 och SAOL 13.

Många av uppslagsorden behövde dock normaliseras, främst på grund av den stora stavningsreformen 1906 (alltså snart efter att upplaga 7 kom ut), men även till följd av senare ortografiska förändringar. I utdraget ur SAOL 1 i Figur 1 gäller det t.ex. **godhjertad** (>**godhjärtad**). Här liksom i andra avseenden är det SAOL 13 som utgör normen. För gammalstavade ord i de tidigare upplagorna som utmönstrats, t.ex. i upplaga 8, ansattes en tänkt nutida stavning, som i allmänhet gav sig naturligt. Exempelvis fick ordet *Afdånande* normalformen *avdånande*.

I de tryckta upplagorna 1–11 tillämpades en nischalfabetisk modell och i upplagorna 12–13 (och 14) en glattalfabetisk (se Figur 1). Dessa modeller återspeglade ett etymologiserande respektive ett rent formellt sätt att betrakta ordförrådet. I princip gjordes två likstavade ord med olika etymologier till två uppslagsord i de elva första upplagorna, och likstavade ord med samma etymologi presenterades som ett enda lemma, ofta även om de tillhörde olika

ordklasser. I upplaga 12 och senare hölls likstavade ord med samma böjning och uttal samman som ett uppslagsord, och likstavade ord med olika böjning eller uttal presenterades alltid som två uppslagsord. Den teoretiska grunden var den s.k. *lemma-lexemmodellen* (se t.ex. Allén 1999).

Skillnaden mellan de båda modellerna kan åskådliggöras med hjälp av ett enkelt exempel, ordet *ljus*. Ordet är både substantiv och adjektiv och behandlas därför som två lemman i SAOL 12–14. Men oavsett ordklass är det förstås fråga om samma etymologi, och *ljus* tas därför upp som *ett* lemma i tidigare upplagor av SAOL (se Figur 2).

<p>ljus 1 -et; pl. = s.; föra ngn bakom l-et lura; klassens l. 2 -t -are adj.; stå i l-an låga brinna klart el. för fullt; till l-an dag -arm s. -behandling -beige s. o. adj. -beständig -bild -blink ~ar -blixt -blond -blå -blått s. -bringare s. -bringerska s. -brun -brunn -brytning -båge -bågs svetsning äv. -båge- -dränkt adj. -dunkel s. -effekt -fenomen -fest -filter</p>	<p>¹ljus s. -et; pl. = • föra ngn bakom ljuset lura; klassens l. ²ljus adj. -t -a • stå i ljusan låga brinna klart el. för fullt; mitt på ljusan dag ljusan se under ²ljus ljusarm s. ljusbehandling s. ¹ljusbeige adj. ²ljusbeige adj.</p>
--	--

Figur 2: Artikeln *ljus* i SAOL 9 och SAOL 13.

Därmed uppstår ett problem: ska *ljus* tas upp som ett eller två uppslagsord i SAOLhist, som ska täcka alla upplagor? Även här bestämde vi oss för att låta SAOL 13 utgöra normen, och därmed står *ljus* som två lemman i SAOLhist. Se Figur 3.

Normaliserad ordform	SAOL 8 (1923)	SAOL 9 (1950)	SAOL 10 (1973)	SAOL 11 (1986)	SAOL 12 (1998)	SAOL 13 (2006)
<i>ljus</i> , adj	ljus2 (a.)	ljus2 (adj.)	ljus2 (adj.)	ljus2 (adj.)	ljus (adj.)	ljus (adj.)
<i>ljus</i> , subst	ljus1 (s.)	ljus1 (s.)	ljus1 (s.)	ljus1 (s.)	ljus (s.)	ljus (s.)

Figur 3: Behandlingen av ordet *ljus* i SAOLhist.

SAOLhist-databasen kan presenteras som en matris med över 200 000 rader (en rad för varje lemma som förekommit i minst en SAOL-upplaga) och elva kolumner (nio SAOL-upplagor, Dalins ordbok och vänsterkolumnen med de normaliserade uppslagsformerna). Ett valfritt urval av kolumnerna, med undantag för vänsterkolumnen, kan undertryckas. Varje ruta är länkad till

motsvarande faksimilsida i den aktuella ordboken, där hela artikeln går att läsa. Tack vare särskilda sökverktyg kan man lätt få fram när ett ord första eller sista gången togs med i SAOL, vilka ord som tillkommit respektive försvunnit mellan två upplagor etc. Man kan söka såväl baklänges som framlänges, vilket ger möjlighet till intressanta morfologiska studier.

3. Datalogiska aspekter

Uppbyggnaden av SAOLhist-systemet finansierades med projektmedel från Svenska Akademien, utan några åtaganden om fortsatt finansiering av underhåll och drift. Vid designen av databasen och gränssnittet var det bl.a. av den anledningen nödvändigt att inta en minimalistisk hållning och välja lösningar som inte medförde licenskostnader eller andra fasta kostnader.

Målet var att med hjälp av ett självförklarande och flexibelt webbgränssnitt, och med respekt för användarens integritet, göra det historiska materialet tillgängligt för både lekmän och forskare. Tack vare samarbetet med Dansk Sprognævn under år 2014 kunde vi vidareutveckla funktionaliteten och konfigurationsmöjligheterna så att samma programpaket kan användas mot olika databaser.

Som databasmotor används MySQL/MariaDB och programmen är skrivna i scriptspråket php. Databasstrukturen är väldigt enkel, för varje ordlista finns en innehållsförteckning med sidreferenser och en tabell med grundinformation. Denna tabell innehåller kolumnerna ”normaliserad ordform”, ”ursprunglig ordform”, ”ordklass”, ”sidhänvisning” och ”ordklassgrupp”. Istället för ordklass och ordklassgrupp kan valfritt annat klassificeringsbegrepp användas – i vår interna miljö använder vi även böjningsinformation i klassificeringen för att åstadkomma homografseparering. Utifrån dessa tabeller genereras tre aggregerade tabeller:

- en med alla ordformer som ska vara sökbara (med kolumnerna: ordform, normaliserad form, bok)
- en med länkar och länktext per normaliserad form och bok (hur länkar och länktext ska genereras specificeras separat för varje boktabell)

- en med länkar och länktext per normaliserad form, ordklassgrupp och bok (hur länkar och länktext ska genereras specificeras separat för varje boktabell)

Användargränssnittet består av en välkomstsida där man väljer vilka böcker som ska ingå i bearbetningen och en huvudsida för sökning och tabellvisning. Detaljinformation presenteras huvudsakligen i pop-upfönster. Avancerade sökparametrar kan visas eller döljas allt efter behov. Vid en enkel sökning behöver man endast fylla i ett ord, med eller utan jokertecken (wildcards).

4. Exempel på sökningar

För att sökresultat i SAOLhist ska vara intressanta från svenska språkets (och inte bara från SAOL:s) synpunkt, krävs förstås att urvalet ur det samtida ordförrådet i var och en av SAOL-upplagorna är någorlunda representativt. Man kan räkna med att i det närmaste alla viktiga enkla (icke-sammansatta och icke-avledda) ord är med, åtminstone från och med upplaga 7.¹ Man kan också räkna med att urvalet av sammansättningar och avledningar till stor del styrs av frekvenskriterier, särskilt i upplagorna 11–13 som är korpusbaserade, men troligen också i tidigare upplagor. I synnerhet de sju nyaste SAOL-upplagorna i SAOLhist-databasen bör därför kunna säga något intressant om det svenska ordförrådets utveckling under mer än 100 år.

Vi ska se på två olika sökningar, en morfologisk och en semantisk. Vid samtliga sökningar inkluderar vi även den internt tillgängliga SAOL 14. Av utrymmesskäl tas i allmänhet bara en delmängd av sökresultaten med. Den första morfologiska sökningen är tydligt relaterad till den aktuella samhällsdebatten. Eftersom klimatfrågorna har kommit i centrum under de senaste decennierna, kan man vänta sig en ökning av antalet sammansättningar med *klimat*- i de senaste SAOL-upplagorna. Så är också i hög grad fallet (se Figur 4). Intressant nog får man inte alls samma utslag om man i stället söker på efterledssammansättningar med *-klimat* (se Figur 5).

¹ De sex första upplagorna hade en viss puristisk prägel, varför många lånord som var fullt brukliga inte kom med. Se t.ex. Malmgren (2002).

SAOL 12 (1998)	SAOL 13 (2006)	SAOL 14 (2015)
klimat (s.)	klimat (s.)	klim·at (s. +et; pl. +)
klimatanläggning (s.)	klimatanläggning (s.)	klim·at/an·lägg·ning (s. +en +ar)
		klim·at/an·passa (v. +de +t)
		klim·at/an·pass·ning (s. +en +ar)
		klim·at/av·tal (s. +et; pl. +)
		klim·at/be·red·ning (s. +en +ar)
		klim·at/bov (s. +en +ar)
klimatbälte (s.)	klimatbälte (s.)	klim·at·bälte (s. +t +n)
		klim·at/de·batt (s. +en +er)

Figur 4: Sökresultat för söksträngen *klimat*%.

SAOL 12 (1998)	SAOL 13 (2006)	SAOL 14 (2015)
	affärsklimat	af·färs klim·at
	arbetsklimat	arbets klim·at
bioklimat	bioklimat	bio klim·at
	börsklimat	
debattklimat	debattklimat	de·batt klim·at
		driv·hus klim·at
ekvatorialklimat	ekvatorialklimat	ekv·at·ori·al klim·at
fastlandsklimat	fastlandsklimat	fast·lands klim·at
	företagarklimat	före·tag·ar klim·at
företagsklimat	företagsklimat	före·tags klim·at
förhandlingsklimat	förhandlingsklimat	för·handl·ings klim·at

Fig. 5. Sökresultat för söksträngen *%klimat*.

Trots att betydelsebeskrivningarna i SAOL ofta är mycket kortfattade, går det faktiskt också att göra vissa semantiska studier med hjälp av SAOLhist.

Ett exempel erbjuder halvsynonymerna *doft* och *lukt*. För många språkbrukare står *doft* alltid för en angenäm och *lukt* för en oangenäm förnimmelse. Denna intuition kan kontrolleras mot efterledssammansättningarna med *-doft* respektive *-lukt* i SAOLhist. Här tänker vi oss närmast en synkron undersökning och begränsar oss till de senaste upplagorna (se Figur 6 och 7).

En snabb blick räcker för att konstatera att språkbrukarnas intuitioner i det närmaste stämmer. Det enda som egentligen stör bilden är ordet *spridoft*, som dock möjligen uppfattas som positivt laddat av vissa språkbrukare.

skunk lukt		
skur lukt	skur lukt	
skvattram[s] lukt		
snusk lukt	snusk lukt	snusk lukt
sprit lukt		
späck lukt		
stall lukt	stall lukt	
stek lukt	stek lukt	

Figur 6: Sökresultat med exempel från upplaga 9, 10 och 11 för söksträngen *%lukt*.

rökelse doft	rökelse doft	rökelse doft
schersmin doft	schersmin doft	
skur doft		
skvattram[s] doft	skvattram[s] doft	skvattram [s] doft
smultron doft		
sprit doft		sprit doft
stek doft	stek doft	
syren doft	syren doft	syren doft

Figur 7: Sökresultat med exempel från upplaga 9, 10 och 11 för söksträngen *%doft*.

5. Slutord

I artikeln har vi velat visa upp SAOLhist som en resurs för olika typer av användare, från forskare till allmänt språkintresserade. De båda exemplen på sökningar är i och för sig av relativt avancerat slag, och betydligt enklare sökningar är givetvis också möjliga. Möjligheterna till trunkerade sökningar gör också SAOLhist användbar för exempelvis korsordslösare och Word-Feud-spelare. En av de vanligaste frågorna som ställs till SAOL-redaktionen handlar om i vilken upplaga ett visst ord kom in i ordlistan, och sådana frågor kan man lätt få svar på i SAOLhist. Det ska slutligen påpekas att den indexering av SAOLhist som är tillgänglig på nätet fortfarande inte är hundra procentigt korrekt; påpekanden om kvarvarande felaktigheter tas tacksamt emot. Tack vare de lätt åtkomliga faksimilsidorna finns dock alltid originaltexterna till hands.

Litteratur

Ordböcker

Dalin, A.F. (1850–55): *Ordbok öfver svenska språket*. Stockholm.

RO = *Retskrivningsordbogen*, udg. af Dansk Sprognævn. Udg. 1–4. 1955–2012. København.

SAOL 1, 6, 7, ..., 14 = *Svenska Akademiens ordlista*, uppl. 1, 6, 7, ..., 14. Stockholm 1874–2015.

Övrig litteratur

Allén, Sture (1981): The lemma-lexeme model of the Swedish Lexical Database. I: Ralph, Bo (red.). *Modersmålet i fäderneslandet. Ett urval uppsatser under fyrtio år av Sture Allén*. Meijerbergs arkiv för svensk ordforskning 25. Göteborg 1999.

Holmer, Louise (2012): SAOLHist – alla upplagor av SAOL i en och samma databas. I: Eaker, B., L. Larsson & A. Mattisson (red.), *Nordiska studier i lexikografi 11. Rapport från Konferensen om lexikografi i Norden, Lund 24–27 maj 2011*. Skrifter utgivna av Nordiska föreningen för lexikografi 12. Oslo, 287-295.

Malmgren, Sven-Göran (2002): Normering i Svenska Akademiens ordlista 1874–1950: principer och resultat. I: *LexicoNordica* 9, 5-20.

Malmgren, Sven-Göran (2014): Svenska Akademiens ordlista genom 140 år: mot fjortonde upplagan. I: *LexicoNordica* 21, 81-98.

Louise Holmer
bitr. forsk., doktorand
Inst. för svenska språket
Göteborgs universitet
Box 200
SE-405 30 Göteborg
louise.holmer@svenska.gu.se

Sven-Göran Malmgren
professor emeritus
Inst. för svenska språket
Göteborgs universitet
Box 200
SE-405 30 Göteborg
malmgren@svenska.gu.se

Monica von Martens
systemutvecklare
Inst. för svenska språket
Göteborgs universitet
Box 200
SE-405 30 Göteborg
monica.von.martens@gu.se

Från ”tandborstord” till ”memilord” – om nyord och deras belägg

Håkan Jansson

This presentation describes an inquiry into recent neologisms in Swedish as presented by *Språkrådet* ('Language Council of Sweden') and *Språktidningen* ('the Language magazine'). Lists of neologisms from 2004 to 2014 has been excerpted and studied together with corresponding language examples. The study shows that the share of in corpora un-attested or only meta use attested neologisms has risen to around one fourth of the total in recent years.

1. Inledning

Ord som används i kontexter som sällan förekommer i skrift – s.k. ”tandborstord” – har länge utgjort ett problem i lexikografernas lemmaurvalsprocess. Idag finns det en annan grupp ord – ”memilord”¹ – som skulle kunna vara ett exempel på ett motsatt problem; de uppmärksammas möjligen i högre grad än de används.

Under senare tid har det blivit vanligt att det gångna årets nyord presenteras vid varje års början. I Sverige är det Institutet för språk och folkminnen som sammanställer nyorden, och de publiceras sedan i *Språktidningen* och i övrig svensk press. Att ett språks ordförråd förändras t.ex. genom att nya ord tas i bruk är en självklarhet; däremot är det långt ifrån självklart vilka dessa nya ord är, och i vilken grad en nyordslista speglar de nya ordens spridning och bruk.

Detta bidrag avser undersöka ett urval av det senaste decenniets svenska nyord så som de har presenterats av Institutet för språk och folkminnen och *Språktidningen*. De utvalda ordens vanlighet och bruk undersöks sedan med hjälp av korpusverket *Korp* och sökmotorn *Google*. Särskild

¹ *Memilord* är ord som uppmärksammas i högre grad än de används. Av *memil* [MEdelålders Man I Lycra], ”medelålders man klädd i cykeldräkt av lycra, som cyklar hänsynslöst; kallas även *ninjacyklist* och *härskarcyklist*”. (Språkrådets nyordslista 2012).

uppmärksamhet ägnas åt frågan huruvida förekomsterna av de undersökta orden är metaspråkliga eller inte; många nyord blir nämligen omskrivna snarare än använda i vanlig kommunikativ funktion.

1.1. Intresset för nyord

Inom den nordiska språkvården har det sedan en längre tid funnits ett intresse för nyord. Så påpekar t.ex. Reuter (2009:200) att man redan vid det första samnordiska språkvårdsmötet 1954 ansett att det var viktigt att uppmärksamma nyordsarbetet för att undvika att onödiga skillnader mellan språken uppstår vare sig i fackspråk eller i allmänspråk:

... av särskilt intresse är att uppmärksamma de nya uttryck av allmän-
nare natur som ständigt dyker upp.

[Ur protokollet från första nordiska språkmötet 1954]

Språknämndernas arbete var kontinuerligt, men det berörde mest språkvården internt och i mindre grad allmänheten. Enligt Agazzi (2014) publicerades visserligen årligen nyord i de nordiska språknämndernas gemensamma årsskrift *Nordiska språkfrågor* (1956–1968), men att avsikten främst var att skapa koordination mellan de nordiska språken framgår bl.a. av Sudmann (1970). Under 1980-talet kom dock nyorden att uppmärksammas mera offentligt i samband med att det gavs ut nyordböcker i alla de Skandinaviska länderna (*Nyord i norsk: 1945–1975*, 1982; *Nye ord i dansk 1955–75*, 1984; *Nyord i svenskan från 40-tal till 80-tal*, 1986). Utan att spekulera alltför mycket om kopplingen till framstegen inom korpuslingvistik, kan man konstatera att den svenska nyordssamlingen kom ut samma år som *Svensk Ordbok* (SOB), som ett år före *Collins COBUILD* var den första ordboken att i grunden vara byggd på korpusmaterial.

I samband med en ny våg av skandinaviska nyordssamlingar (Jarvad 1999, Moberg 2000a, Guttu & Wangenstein 2012) började språknämnderna också publicera upplysningar om nyorden mera tillgängligt. I Sverige skrevs det oregelbundet om nyord i *Språkvård* under tiden 1987–1999 (Moberg 2000b), men sedan 2000 har årets utvalda nyord publicerats i *Språkvård* och dess efterföljare *Språktidningen*. Under senare tid har det blivit vanligt att det gångna årets nyord presenteras vid varje års början. I Norge lyfter Språk-

rådet, förutom listan över årets ord, fram ett av dem som årets ord. Principerna bakom valet orden på den norska listan diskuteras utförligt av Våge & Andersen (2014), som dessutom ger en utblick över nyordsarbetet i ett tiotal andra länder. 2013 var årets ord i Norge enligt Språkrådet *sakte-tv*. Språkrådets ord får å andra sidan konkurrens av andra privata utnämningar som kan gå på tvärs mot språkvårdens rekommendationer. Tidningen VG utkordade i en spektakulärt uppslagen artikel på egen hand *selfie* till årets nyord 2013.

I samband med att lanseringen av nyorden blir mer spektakulär kan det vara på sin plats att undersöka i vilken grad de faktiskt återspeglar ett allmänt bruk – och om de kan förväntas vara mer än dagsländor. Just att tillfälligheter i bruket kan ligga bakom de presenterade nyorden har påpekats av Heidemann Andersen et al. (2015) i samband med att de presenterade 2014 års danska nyord. De använder ordet *vuvuzela* för att illustrera hur bruket av ett uttryck kan ha vissa toppar (här sommaren 2010 och årsskiftet 2010-2011) för att sedan knappt användas alls.

2. Undersökningsmaterial och utgångspunkter

Utgångspunkten för de reflektioner som presenteras här, är en jämförelse mellan de nyord som lanserats av *Språkrådet* och *Språktidningen* 2004 till 2014² och språkbruket, som det kan studeras i relevanta referenskorpusar. Här valdes två större korpusar som sattes samman ur de texter som finns tillgängliga i Korp – korpusverktyget som tillhandahålls genom Språkbanken vid Institutionen för svenska språket vid Göteborgs Universitet. För ytterligare upplysningar om Korp se t.ex. Borin et al. (2012). Den ena korpusen har hämtats från Språkbankens rikssvenska nyhetstextkorpusar (Göteborgsposten [GP] 1994, 2001–2013, GP:s helgbilaga Två dagar; Press 65, 76, 95, 96, 97, 98; Webbnyheter 2001–2013, Dagens Nyheter 1987; 8 sidor; Ordat.). En mindre mängd äldre texter har tagits med för att kontrollera om orden tidigare har varit i bruk. Totalt omfattar denna korpus – hädanefter *Presstext (P)* – 575 miljoner löpord. För att även kunna beakta ett huvudsakligen informellt språkbruk används Språkbankens bloggmix från åren

² På grund av tidsbrist har 2009 tyvärr inte kunnat analyseras.

1998–2014 – hädanefter *Bloggtexter* (B) – som omfattar 505 miljoner löpord.

Förutom jämförelsen med korpusar är det möjligt att indirekt se nyorden i relation till erfarna lexikografers bedömningar i och med att en ny utgåva – den fjortonde – av SAOL utkom i april 2015. Således undersöks även i vilken omfattning de lanserade nyorden förekommer i SAOL (2015).

3. Nyordens relevans i förhållande till korpusmaterialet

Nyordens relevans i förhållande till det bruk som noteras kan bedömas på flera olika sätt. Man dels undersöka hur många nyord som helt saknar belägg i dessa förhållandevis stora korpusar, dels notera andelen metaspråkliga belägg – d.v.s. belägg som uppmärksammar orden utifrån den nyutkomna nyordslistan.

3.1. Nyord utan belägg i korpusarna

Nyorden utan helt belägg i vare sig press- eller bloggkorpus är förhållandevis få. Vidare refereras till flertalet av dem med belägg i andra tidningar än de som ingår i korpusen. I följande uppställning anges hur många av nyorden varje år som saknade korpusbelägg, och några beläggslösa ord ges som exempel:

2004	4 av 58	blingblingare, kyrktrappsbröllop, lokator, luvunge
2005	14 av 62	hotelljournalistik, id-sprej, koscheria, industrisafari
2006	13 av 59	minnesprick, stadsglesning, stadsspret, timellare ³
2007	4 av 59	interaktiv skrivtavla, munhota, wikipedisk
2008	2 av 31	ljudaffisch, mobillångfilm
2010	8 av 51	filvärd, kringtidskort, solhybrid, valpromenera
2011	1 av 73	döda vinkeln-varnare
2012	2 av 42	mjuk betalvägg, mossgraffiti
2013	6 av 41	dygnis, global hektar, linjär tv, rutkod, tvåhandsbeslut
2014	14 av 41	attefallshus ⁴ , klickfiske, kringis, kärtrorpa ⁵

³ 'Hemmasnickare' efter tv-snickaren Martin Timell.

Vissa av orden utan belägg i korpusarna har inte heller särskilt många belägg på Google. Så har t.ex. "blingblingare" färre än 20 belägg vid en sökning 27 september 2015. Av dessa belägg var ordet i många av fallen sedan heller inte använt så som det definierats av Språkrådet – 'person som bär blingbling' – utan helt enkelt som synonym till blingbling: "Matchar man med några grymma blingblingare och häftiga skor (som på bilderna) så har man en snygga avslutningsoutfit som kan hänga med hela dagen."⁶ Ordet lanserades 2004. Således finns det knappt ett relevant belägg per år indexerat på Google i förhållande till tiden sedan lanseringen. Andra – nyare – ord kommer möjligen att om några år märkas i korpusarna med tanke på att de har en rikligare representation på Google. Så ger t.ex. "kringis" 72 träffar inom ett år från att ordet har lanserats.⁷

3.2. Nyord med hög andel metaspråkliga belägg

Även om nyorden som helt saknar belägg i korpusarna är förhållandevis få, är desto fler ord huvudsakligen belagda med texter som rapporterar om nyordslistan, i synnerhet för senare delen av perioden. Därför finns det skäl att titta något närmare på dem. Här nedan presenteras ord från presskorpusen (eller om belägg saknas där, från bloggkorpusen – B) med antalet metabelägg i förhållande till de totala beläggen:

2004:

brattig 4/5; curlingförälder 19/162; flexitarian 1/1; kotlettfrilla 3/8; servicebarn 4/4; webblogg 8/23; vuxenvälling 2/7; örådisera 1/2.

⁴ 'Hus som kan byggas utan bygglovsprövning', efter ansvarig minister Stefan Attefall.

⁵ Att 'göra motstånd mot nynazism' efter Stockholmsförorten Kärrtorp där man konfronterade nynazisterna.

⁶ Google 27/9 2015: <tsundin.blogg.se/2010/march/schools-out-soon.html>.

⁷ Google 27/9 2015: https://www.google.se/search?num=100&lr=lang_sv&cr=countrySE&safe=images&hl=sv&biw=1920&bih=916&tbs=ctr%3AcountrySE%2C%3AAlang_1sv&q=%22kringis%22&oq=%22kringis%22&gs_l=serp.12...32799.36864.0.43249.2.2.0.0.0.110.192.1j1.2.0....0...1c.1.64.serp..1.1.109.fTiaM0iNPLw

2005:

flyttstajning/flyttstajling/flyttstyling 1/1; gala in 1/1; glaskulefolket 5/5; poddare 2/4.

2006:

flexidaritet 1/1; fulbryt 1/1 (B); Odellplatta 2/10; sverka 2/6.

2007:

agflation 2/8; damp 2/3; nyhetsbok 1/1 (B); reinfeldtare 2/12.

2008:

ADV 2/2; ansiktsspårning 2/2; bekymringssamtal 3/14; blåstråle 1/1; bullerbysyndromet 1/1; celebritariat 1/2; finanssmälta 5/5; gatustickning 1/1; kalsongbadare 1/3; klimatism 2/5; skynka 5/5; ödleplåster 1/2.

2010:

askbränd 1/4; askstoppad 1/7; asktåg 1/2; askänka/-änkling, vulkanänka/-änkling 4/5; bjästa 2/17 (B); danseoke 1/2 (B); Facebookfest 1/1 (B); guldomat 1/1 (B); inaskad 4/7; jobbstopparpolitik 1/1 (B); kaffeflicka 4/6; kalkstensångare 1/2; kylkrage 1/2 (B); morotsaktivism 1/1; morotsmob 2/4; näringslots 1/1; spotifiera 1/7; wikiläcka 1/8; vulkanresa 2/3.

2011:

ac-förkylning 1/1; ac-snuva 1/1; airbaghjälm 1/1; airbagjacka 1/1; appa 1/2; attitydinkontinens 2/3; Bamseteorem 1/2; brorsantrick 1/5; brännskräp 2/2; fastlans 1/2; flipperförälder 1/2; foliehatt 11/34; fulparkerare 2/2; förlåtelseurné 2/12; förväntis 2/2; gps-väst 1/6; jasminmöte 1/3; kapselbryggare 1/4; knarkometer 2/2; loba 1/1; lunchdisco 1/8; mobildagis 2/2; molekylärgastronomi 1/9; möbelhund 1/1; nysare 1/1; otrohetsdejting 1/1; plankning 1/1; post-it-krig 1/7; restdejting 1/1; retronym 1/8; seismisk 1/7; skräpbot 1/6; skämsfilter 1/9; slöjböter 1/1; sms-livräddning 1/1; surdegfaktorn 2/4; tasigförsamhet 1/1; terja 5/6; tjejsamla 7/8; tonårsskrämma 2/2; trollfilter 2/2; trädmod 1/5; ugpling 1/1; ungdomsskrämma 1/1; vattkoppsgodis 2/2; åsiktstaliban 2/3.

2012:

conversesjukan 1/1; drinkorexi 3/5; emoji 1/4; eurobävning 2/4; grexit 2/15; henifiera 7/7; klämspärr 3/3; kramtjuv 1/4; kötrymden 4/5; livslogga 1/4; läshund 2/2; memil 15/19; märk-dna 1/1; nomofobi 3/3; pekskärmsvantar/surfvantar/touchvantar/mobilvantar 2/6; queerpolska 1/4; ryggplankning 2/2; ståhjuling 2/5; Tintingate 1/3; tårtgate 2/4; utvigning 1/1; zlatanera 20/40; underklassafari 2/10; överklassafari 2/26.

2013:

betalskugga 1/1; carpa 1/1; funktionell dumhet 1/1 (B); gubbploga 1/1 (B); hämndporr 1/2 (B); kjolprotest 1/1 (B); klicktivism 1/1 (B); snippgympa 1/6 (B).

2014:

fotobomba 1/11 (B); krislåda 1/1 (B).

Vikten av att ha två olika typer av korpusar blir uppenbar vid betraktandet av ord som *dampa* 'vara bråkig och störande; få utbrott' (2007), *terja* 'manipulera ett foto i bedrägligt syfte' (2011) och *carpa* 'passa på att njuta' (2013). Alla tre orden kännetecknas av metabruk i sammanställningen ovan, I blogg-korpusen noteras å andra sidan att *dampa* har 751 belägg, *terja* 125 belägg och *carpa* 49 belägg, och bara ett av dessa är metaspråkligt.

Ett av orden ovan, *tjejsamla* (2011), inkluderades som nyord på grundval av *ett* (!) belägg i ett kåseri. Som följd av nyordsstatusen kunde detta enda belägg skapa en debatt i media och elektronisk offentlighet, som ledde till att ordets spridning ökade väsentligt.

4. Nyorden och SAOL

Förra avsnittets relativt omfattande exempel på nyord som tycks ha liten och huvudsakligen metaspråklig spridning, kan motivera en liten studie i relationen mellan de lanserade nyorden och SAOL 14, för att se om dessa ord haft någon påverkan på lexikografen.

Till att börja med ska det konstateras att nyordslistorna har gått igenom av SAOL-redaktionen som ett led i lemmaurvalsprocessen för den nya ordlistan. Av följande uppräknig av antal medtagna ord (med totalt antal nyord

inom parentes) per år, framgår att något under hälften av orden tar plats i den nya upplagan av SAOL: 2004 19 (58); 2005 32 (62), 2006 31(59), 2007 30 (59), 2008 16 (31), 2010 14 (51), 2011 20 (73), 2012 13 (42), 2013 14 (41), 2014 6 (41). Av uppräknigen framgår också att andelen inkluderade ord efter 2010 minskar från ca en tredjedel till ca en sjundedel 2014. Som framgår av nedanstående diskussion om ordens beläggsstatus, är det just efter 2010 som det dessutom sker en ökning av ord med stor andel metaspråkliga belägg.

Nyord utan belägg i press/bloggkorpusarna som har kommit med i SAOL är mycket få: *solhybrid* (2010), *mossgraffiti* (2012), *smartplåster* (2013) och *selfiepinne* (2014). Även om SAOL-redaktionen naturligtvis har haft tillgång till nyordslistorna tycks det inte ha påverkat deras bedömning i dessa fall. Ord med få träffar har inte i någon nämnvärd grad tagits med. Bland undantagen finns ord som *smartplåster* och *selfiepinne* med ca 370 respektive ca 310 träffar på Google 28 september 2015, som har bedömts intressanta trots att de saknar träffar i korpusarna.

Ord med hög andel metaspråkliga belägg är betydligt mer välrepresenterade i SAOL 14: *curlingförälder*, *vuxenvälling* (2004), *flyttstyling* (2005), *bekymringssamtal*, *finanssmälta*, *kalsongbadare* (2008), *wikiläcka* (2010), *airbaghjälme*, *appa*, *foliehatt*, *gps-väst*, *möbelhund*, *retronym*, *trollfilter*, *träd-mord* (2011), *emoji*, *grexit*, *köttrymden*, *memil*, *pekskärmsvantar*, *överklassafari*, *underklassafari* (2012) samt *betalskugga* (2013). Vissa av dessa ord har relativt höga frekvenser på Google och som i fallet *bekymringssamtal* dessutom en koppling till likabetydande danska och norska ord. I andra fall som *möbelhund* och *trollfilter* med ett respektive fem belägg, är samtliga metaspråkliga. Det är dessutom så att Google ger ett bedrägligt antal träffar eftersom *möbelhund* finns som tyskt ord och *trollfilter* som engelskt och tyskt. Vid en enkel googling (28/9-15) erhålls 26 500 resp. 4 600 träffar, men om googlingen begränsas till svenska ord minskar träffarna till 50 resp. 95. Av de 95 träffarna för *trollfilter* är 49 metaspråkliga och träffar för 2014 saknas. *Möbelhund* har 3 träffar 2014 och av de 50 träffarna är 20 metaspråkliga. Båda orden har dessutom ett mindre antal träffar som är helt irrelevanta, som att de kopplas till sökrobotar eller att de inte har den här avsedda betydelsen. Det som vid ett hastigt påseende kan tyckas vara ett stort antal belägg, är således inte alls lika övertygande vid en närmre undersökning.

5. Avslutning

Det ökande intresset för nyord som har kunnat konstateras hos såväl språknämnder som medier, borde med utgångspunkt i det som har rapporterats här ge anledning till två reflektioner:

1. Språknämndernas nyordsarbete har fått en inriktning som skiljer sig från den som var gällande då de började sitt internordiska samarbete. Då fanns en tydlig vilja till språkplanering såtillvida att man ville undvika att det uppstod onödiga skillnader mellan språken. En samordning av nyorden sågs som viktigt. Idag tycks strävan att dokumentera vara helt överordnad nordisk samordning.

2. Vi ser här att viljan att dokumentera och snabbt föra ut iakttagelserna leder till att ord med liten frekvens får en ”oavsiktlig (?)” lansering. 13% av nyorden saknade belägg i någon av korpusarna och ytterligare 25% hade en stor andel metaspråkliga belägg.

Med ovanstående i minnet, kanske denna undersökning har väckt frågan, om inte en återgång till den modell för nyordsarbete som språknämnderna upphavligen använde vore önskvärd.

Litteratur

Ordböcker

COBUILD (1987): = *Collins COBUILD English Language Dictionary*.

Editor in Chief: John Sinclair, Managing Editor: Patrick Hanks.
London/Glasgow: Collins.

Guttu, Tor & Boye Wangensteen (2012): *Nyord i norsk*. Oslo: Kunnskapsforlaget.

Jarvad, Pia (1999): *Nye Ord. Ordbog over nye ord i dansk 1955-1998*. København: Gyldendal.

Moberg, Lena (2000a): *Nyordsboken. Med 2000 nya ord in i 2000-talet*. Stockholm: Svenska språknämnden.

Norsk språkråd (1982): *Nyord i norsk: 1945–1975*, Bergen: Univ.-forl.

Nyord i svenskan från 40-tal till 80-tal (1986). Solna: Esselte studium.

- Riber Petersen, Pia [=Jarvad, Pia] & Jørgen Eriksen (1984): *Nye ord i dansk 1955-75* [Dansk Sprognævns skrifter 11]. Köpenhamn: Gyldendal.
- SAOL = *Svenska Akademiens ordlista över svenska språket* (2015): 14. uppl. Stockholm: Svenska Akademien/Norstedts akademiska förlag [distributör].
- SOB = *Svensk ordbok* (1986). Solna: Esselte studium.

Övrig litteratur

- Agazzi, Birgitta (2014): Nordiskt nyordssamarbete i går, i dag och i morgon. I: Heidemann Andersen, Margrethe, Pia Jarvad & Jørgen Nørby Jensen (red.): *Neologismer. Dansk Sprognævns 2. seminar om nye ord*. Frederiksberg: Dansk Sprognævn.
- Borin, Lars, Markus Forsberg & Johan Roxendal (2012): Korp – the corpus infrastructure of Språkbanken. I: *Proceedings of the eighth international conference on Language Resources and Evaluation*. Istanbul. [tillgänglig elektroniskt: <www.lrecconf.org/proceedings/lrec2012/pdf/248_Paper.pdf>].
- Heidemann Andersen, Margrethe, Pia Jarvad & Jørgen Nørby Jensen (2015): *Normcore, spornoseksuel og sorgsnylter – ord og udtryk fra året der gik. Dansk Sprognævns nyordsliste 2014*.
- Moberg, Lena (2000b): Några nyare ord i svenskan. *Språkvård* 1, 16–24. *Nordiska språkfrågor* [da. *Nordiske sprogproblemer; no. Nordiske språkspørsmål*] 1956–1968. [Skrifter utgivna av Nämnden för svensk språkvård]. Stockholm: Läromedelsförl. (Sv. bokförl.).
- Reuter, Michael (2009): Nordiskt språksamarbete via de nordiska språkmötena sedan 1954 – har de varit till nytta för språkvård och språkforskning. I: *Språk i Norden 2009: Tema: Språkpolitik och språkattityder*. Oslo: Språknämnderna i Norden, 199–214.
- Sudmann, Arnulv (1970): Nordisk språksamarbeid. I: *Språk i Norden 1970*. Stockholm: Nämnden för svensk språkvård/Läromedelsförlagen Svenska Bokförlaget, 66–106.
- Våge, Ole Kristian & Gisle Andersen (2014): Nyord, kriterier og språknormering i en årlig kåring. I: Heidemann Andersen, Margrethe, Pia Jarvad & Jørgen Nørby Jensen (red.) *Neologismer. Dansk Sprognævns 2. seminar om nye ord*. Frederiksberg: Dansk Sprognævn.

Håkan Jansson
Fil.dr. i nordiska språk
Högskolan Väst
SE-461 32 Trollhättan
hakan.jansson@hv.se

Processering af en synonym-komponent i en flersproget ordbog

Halldóra Jónsdóttir & Þórdís Úlfarsdóttir

The ISLEX dictionary combines Icelandic as a source language (SL) and six Nordic languages as target languages (TLs) in a single database. This article describes the process of how Icelandic synonyms can be extracted and grouped together by using the TLs' equivalents. Each word class is processed independently. Selected material, a list of lemmas with their ID number plus their TLs, is exported from the database, going through the languages one at a time. This produces long lists of results which consist of the Icelandic headwords and their equivalents. The lists are then processed in certain ways. In cases where a lemma has more than one sense there is a danger that the boundaries between the senses become blurred so that the words in those groups tend to get mixed up. An example is the Icelandic noun *verslun* which means both 'shop' and 'trade', and is therefore divided into two senses in the dictionary. All the material comes from within the dictionary and no external data is used. It means that all the synonyms obtained in this way are necessarily also lemmas. Despite this we consider the method justifiable because it is quick and simple, and in many instances it gives very good results, i.e. groups of synonyms with varying degrees of exactness.

1. Baggrund

ISLEX er en tværnordisk, flersproget onlineordbog med islandsk som kildesprog og dansk, norsk bokmål, nynorsk, svensk, færøsk og finsk som målsprog. ISLEX er et samarbejdsprojekt mellem seks institutioner i Island, Danmark, Norge, Sverige, Finland og på Færøerne. Disse institutioner er *Stofnun Árna Magnússonar í íslenskum fræðum* (SÁM) i Reykjavík, *Det Danske Sprog- og Litteraturselskab* (DSL) i København, *Institutt for lingvistiske, litterære og estetiske studier* ved Universitetet i Bergen, *Institutionen för svenska språket* ved Universitetet i Göteborg, *Fróðskaparsetur Føroya* i Tórshavn og *Helsinki Universitet*. De to

sidstnævnte institutioner indgik dog i samarbejdet på et senere tidspunkt end de førstnævnte.

ISLEX-ordbogen blev udarbejdet på den måde at den islandske redaktion var ansvarlig for kildeproget samt udformning og udvikling af databasen. Redigeringen af målsprogene derimod blev varetaget af oversætterne i hvert af de øvrige lande. Arbejdet foregik i en webbaseret database specielt lavet til projektet, således at alle landenes redaktioner kunne arbejde samtidig.

Den første version af ISLEX blev åbnet på webben i november 2011 (for målsprogene dansk, norsk og svensk), og den islandsk-færøske ordbog i marts 2015. Den islandsk-finske ordbog er stadig under bearbejdning og bliver åbnet senere.



Figur 1: Landene som deltager i ISLEX-samarbejdet.

ISLEX-ordbogen er den første onlineordbog der omfatter og forbinder flere nordiske sprog. Ordbogen dækker moderne islandsk sprog, med særlig vægt på at repræsentere et mangfoldigt udvalg af kollokationer, idiommer og eksempler med tilhørende oversættelser til målsprogene. ISLEX udnytter de fordele som elektronisk formidling giver mulighed for så som illustrationer og lyd. Der vises bøjningsparadigmer til alle de opslagsord der kan bøjes, via et link til den morfologiske database over islandsk, *Beygingarlýsing íslensks nútímamáls*, som er udarbejdet hos SÁM.

På ordbogens hjemmeside er der forskellige typer af søgemuligheder. Man kan f.eks. vælge at foretage opslag i alle målsprogene samtidigt, for således at få et indblik i det islandske sprogs slægtskab med målsprogene og ligeledes et overblik over indbyrdes forskelle og ligheder mellem sprogene.



Figur 2: Ordbogen kan belyse ligheder og forskelle mellem sprogene. *Ananas* er ens på alle sprogene.

Figur 3: Variation i navnet *Grækenland*.

Fra starten ønskede man at ISLEX blev et bidrag til at styrke de kulturelle forbindelser i Norden og fremme den internordiske sprogforståelse, og det fremgår klart at efter at ordbogen blev åbnet, har den fået stor udbredelse og titusinder af brugere. Desuden gemmer et så omfattende værk på mange muligheder for al slags sprogforskning, og i denne artikel beskrives et nyt projekt: en synonym-komponent for kildesproget, islandsk.

2. Leksikografisk placering

Synonymordbøger i deres enkleste form får almindeligvis ikke særlig stor opmærksomhed indenfor leksikografien, og denne type ordbøger har hidtil ikke nydt stor respekt. Hovedårsagen er sandsynligvis de begrænsede relationer i selve ordforrådet, hvor alle ordene indtager den samme position i en flad struktur. Men selvom synonymordbøgerne måske ikke anses for at være udpræget videnskabelige værker, er de af stor nytte for mange brugere, f.eks. når de skal producere tekster eller løse et krydsord.

Tesaurusser og begrebsordbøger er nært beslægtede med synonymordbøgerne, men de er normalt mere indholdsrige, og indholdet inddeles gerne på en anden måde, oftest hierarkisk under overbegreber. I sådanne ordbøger er der plads til forskellige ord som ikke hører hjemme i synonymordbøger, f.eks. navne på måneder og træsorter. Tesaurus er en særlig ordbog eller

'ordliste hvor ordene er ordnet efter emne og betydning (og altså ikke alfabetisk)' (jf. forklaringen af *thesaurus* i DDO). Eftersom en begrebsordbogs indhold ikke ordnes alfabetisk, er det vigtigt at inkludere et indeks over emnet for at gøre det nemmere for brugerne at finde frem til det som de søger. Dette gælder først og fremmest trykte bøger, og eksempler på dette finder man i Jónsson 2005 og DDB 2014, som begge indeholder et kæmpemæssigt indeks. I de rendyrkede synonymordbøger ordnes ordene derimod alfabetisk hvilket gør det nemmere at finde frem til det man søger.

På islandsk findes der både en stor begrebsordbog (Jónsson 2005) og en synonymordbog (Sigmundsson 2012). Nærværende projekt er således ikke fremkaldt af et akut behov. Projektet er i højere grad en undersøgelse af hvorvidt det er formålstjenligt eller overhovedet muligt at generere synonymgrupper ud af de sprogdata som gemmer sig i ISLEX. På nuværende tidspunkt er det altså ikke meningen at udarbejde en fuldstændig digital synonymordbog, og ej heller en begrebsordbog, men redaktionen anser det som værende et spændende pilotprojekt som muligvis kan danne grundlag for noget større.

3. Materiale til synonymgrupperne

Der er ca. 50.000 opslagsord i ISLEX som dermed bliver kandidater til synonym-processeringen. Hele materialet stammer fra selve databasen og der hentes ingenting udefra, et faktum som selvfølgelig begrænser antallet af enkelte synonymer. På nuværende tidspunkt medtages ikke flerordsenheder som ofte er fyldestgørende synonymer til enkelte ord (jf. dansk *med det samme* og *straks*), og desuden har en del af ordforrådet ikke noget synonym.

Eftersom ISLEX-ordbogen har islandsk som kildesprog er den primære tilgang til ordforrådet gennem et islandsk opslagsord (Figur 4). Men det er også muligt at søge omvendt, dvs. ved at indtaste et ord fra et af målsprogene, hvor der efterfølgende fremstår en liste over de islandske ordbogsartikler som indeholder den ønskede ækvivalent, m.a.o. et udvalg af islandske opslagsord, som matcher søgningen. Figur 5 viser et eksempel på dette.



Figur 4: Et islandsk opslagsord, 'skrímsli', med to danske ækvivalenter, 'uhyre' og 'monster'.

Figur 5: Omvendt søgning, 'monster', på dansk og svensk, med de islandske opslagsord efter flaget.

Det er dog ikke altid muligt at opnå et så omfattende søgeresultat som på Figur 5 når man søger et ord i målsprogene. I nogle tilfælde er der bare ét ord i målsproget imod ét andet ord i kildesproget (*historielærer, onsdag*). Derimod giver visse (ofte ukonkrete) ord i mange tilfælde udmærkede resultater, f.eks. det danske ord *adfærd* som afføder tolv gode islandske synonymer (Figur 6).



Figur 6: Søgeresultat for det danske ord 'adfærd'. Det matcher 12 islandske opslagsord.

Ordgrupperne foroven fremkommer når man søger i ordbogens brugergrænseflade efter et ord på et af målsprogene. Dermed aktiveres en kom-

mando til databasen som besvarer søgningen og frembringer ækvivalent sammen med lemma og lemma-ID som knyttes dertil. Listen for neden viser outputtet fra en direkte SQL-forespørgsel om to bestemte danske ækvivalenter, der er foretaget udenom ordbogens hjemmeside. Rækkefølgen i nedenstående liste er: ækvivalent, ordklasse (no = navneord) og opslagsordet med dets lemma-ID. Dermed opstår der en liste med to danske ord, *adskillelse* og *adspredelse*, hvor de islandske match bliver potentielle synonymer:

adskillelse no	aðgreining: 2328
adskillelse no	aðskilnaður: 2397
adskillelse no	skilnaður: 36414
adskillelse no	sundurgreining: 55696
adskillelse no	viðskilnaður: 45572
adspredelse no	afþreying: 2817
adspredelse no	dægrastytting: 9027
adspredelse no	ðund: 8824
adspredelse no	upplyfting: 43476
adspredelse no	yndisauki: 46608

Hermed får vi en bekræftelse af at denne metode er oplagt til at hente islandske synonymer frem med. Eftersom opslagsordene ikke er systematisk forbundet gennem selve kildesproget, er ækvivalenterne i mange tilfælde den stærkeste forbindelse mellem ordene¹. Databasen indeholder flere tusinde ækvivalenter på hvert af ordbogens seks målsprog, og det er klart at det drejer sig om store mængder af sproglige data som kan anvendes til at generere synonymgrupper, som vises i Figur 5 og 6. Herunder beskrives nærmere den metode som anvendes for at trække synonymerne ud af databasen.

¹ Ordforrådet er i forvejen inddelt i semantiske felter, men inddelingen kan kun i begrænset omfang bruges til at gruppere ordene med.

4. Metode

På samme måde som beskrevet ovenfor gennemgås alle målsprogene, ét for ét, og således knyttes synonymgrupperne for islandsk sammen, med lemmaet som det centrale omdrejningspunkt. Som før nævnt er målsprogene i ISLEX flertallet af de sprog som tales i Norden (dog ikke grønlandsk og samisk). Finsk er det eneste målsprog i databasen som ikke er af samme sprogfamilie som de øvrige sprog. Til trods for dette er resultaterne af processeringen af finsk lige så gode som for de andre sprogs vedkommende, men for enkelhedens skyld fokuserer vi mest på dansk i denne artikel. Processen ved at søge potentielle synonymer sker trinvist, og her beskrives metoden i store træk.

1. Det første trin er en eksport fra databasen: *ækvivalent + ordklasse + opslagsord + lemma-id*. Dette gentages for hver ordklasse for sig og hvert målsprog for sig. Resultatet bliver en liste som den som vises herunder, i dette tilfælde drejer det sig om adjektiver på dansk med et tilsvarende islandsk lemma:

magtløs adj	magnvana: 51413
manglende adj	skertur: 36368
mangelfuld adj	ófullkominn: 31010
metamorfoseret adj	myndbreyttur: 52575
modig adj	skeleggur: 36272
morsom adj	fyndinn: 14317
morsom adj	gamansamur: 14794
morsom adj	skemmtilegur: 36328
motordrevet adj	vélknúinn: 45434
mættet adj	gegnsósa: 15104

2. Næste skridt er at sammenføje lister med resultater fra alle de seks målsprog og ordne dem alfabetisk. På denne måde blandes alle målsprogene sammen i en alfabetisk liste. Den ser sådan ud:

morsk adj	illúðlegur: 22448
morsk adj	yggldur: 46594
morsom adj	broslægur: 7016
morsom adj	fyndinn: 14317
morsom adj	hlægilegur: 20177
morsom adj	kátlegur: 23641
morsom adj	kímilegur: 23868
morsom adj	skemmtilegur: 36328
morsom adj	skondinn: 36792
morsom adj	skoplegur: 36799
morsom adj	sniðugur: 38129
morsom adj	spaugilegur: 38588
mosbevokset adj	mosagróinn: 57790

3. Nu *køres* to små perl-scripter på resultatlisten. Det første script opstiller listens islandske ord (med ID) i samme linje med et komma imellem, efter den danske ækvivalent plus ordklasse.

morsom|adj|broslægur: 7016, fyndinn: 14317, gamansamur: 14794, hlægilegur: 20177, kátlegur: 23641, kímilegur: 23868, kostulegur: 24490, skemmtilegur: 36328, skondinn: 36792, skoplegur: 36799, sniðugur: 38129, spaugilegur: 38588

4. Dernæst bliver det sidste script *kørt* på resultaterne. Scriptet fjerner ækvivalenten (*morsom*) først i linjen og danner par med de ovenstående islandske ord:

broslægur: 7016	fyndinn: 14317
fyndinn: 14317	broslægur: 7016

broslægur: 7016	gamansamur: 14794
gamansamur: 14794	broslægur: 7016
broslægur: 7016	hlægilegur: 20177
hlægilegur: 20177	broslægur: 7016
broslægur: 7016	kátlegur: 23641
kátlegur: 23641	broslægur: 7016

Det er meningen at alle synonymerne fremstår i alle ordbogsartiklerne i listen, og processen skaber en fordoblingseffekt i parrene eftersom alle kombinationer skal parres. I de tilfælde hvor der kun findes to synonyme (*broslægur*, *fyndinn*) bliver linjerne kun to:

broslægur	fyndinn
fyndinn	broslægur

Hvis der derimod er tre synonyme (*broslægur*, *fyndinn*, *gamansamur*) bliver linjerne seks:

broslægur	fyndinn
broslægur	gamansamur
fyndinn	broslægur
fyndinn	gamansamur
gamansamur	broslægur
gamansamur	fyndinn

5. Til slut må man gennemgå hele materialet omhyggeligt fordi metoden indebærer at der altid vil være ord som ikke passer ind. Derefter er materialet klart til indlæsning i databasen eller til et andet formål.

5. Bearbejdning af resultaterne

Der er nu blevet genereret lange lister med materiale til synonymgrupper som er færdige til gennemlæsning. Den mest tidskrævende faktor er at fjerne de ord som ikke passer til synonymlisterne, og som udgør omkring 19% af alle ordene. Det ville faktisk være muligt at opnå en større procentdel af brugbart materiale ved ikke at anvende alle målsprogene i ISLEX som forbindelsesled mellem de islandske ord, noget som også ville medføre at der fremstod færre synonymmer end ellers. Det kræver en vis balancegang at finde frem til den bedste arbejdsmetode.

Under processeringen sker det tit at falske venner dukker op i listerne. I de fleste tilfælde drejer det sig om at ord af samme stamme i nærtbeslægtede sprog ikke har den samme betydning, f.eks. dansk og norsk *rolig* 'stille' og svensk *rolig* 'morsom'. Et andet forstyrrende moment er polysemien blandt ækvivalenterne, f.eks. betyder svensk *affär* både 'butik, forretning' og '(kærligheds)affære', og det danske ord *brud* betyder både 'afbrydelse' og 'en kvinde som skal giftes'. På dette sidste stadie spiller redaktøren hovedrollen for at sikre at kun de relevante ord indgår i listen.

For den ikke-islandske sprogbruger kan det være svært at vurdere hvor gode eller nære synonymerne er, og hvad angår deres valør, stilleje, frekvens osv. ville det uden tvivl være nyttigt at få oplysninger derom. Eftersom det drejer sig om automatisk generering, så er dette ikke så ligetil på nuværende stadie.

6. Konklusion

Ovennævnte projekt kan beskrives som et pilotprojekt, hvor man afprøver en metode ved generering af synonymmer igennem målsprogenes ordforråd i en flersproget ordbog. Resultaterne viser at det er en effektiv og hurtig metode til dannelse af synonymmer. Projektets materiale stammer alene fra indholdet i ISLEX og hvis man ønsker at udvide synonymkomponenten med flere ord, må disse hentes udefra. Desuden skal der gøres specielle tiltag for at det bliver muligt at inkorporere flerordsenheder i synonymgrupperne.

Projektet er i skrivende stund under forberedelse, men planen er at komponenten kommer som en tilføjelse til ISLEX-ordbogen og dets afledte

værker. Et af disse projekter er KATA, en ny islandsk webordbog som er baseret på det islandske materiale i ISLEX. Arbejdet med en ny islandsk-fransk ordbog med arbejdstitlen 'Lexia' er også påbegyndt som et spin-off-projekt af ISLEX. Alle disse værker vil nyde godt af en videre udvikling af ISLEX-ordbogens omfattende materiale.

Litteratur

Ordbøger

- DDB (2014) = Nimb, Sanni (hovedred.): *Den Danske Begrebsordbog*. København: Det Danske Sprog- og Litteraturselskab.
- DDO = Trap-Jensen, Lars et al. (red.): *Den danske ordbog*. København: Det Danske Sprog- og Litteraturselskab. <ordnet.dk/ddo> (september 2015).
- ISLEX = Úlfarsdóttir, Þórdís (red.): *ISLEX-ordbogen*. Reykjavík: Árni Magnússon-instituttet for islandske studier. <islex.is>, <islex.dk> (september 2015).
- Jónsson, Jón Hilmar. (2005): *Stóra orðabókin um íslenska málnotkun*. Reykjavík: Forlagið.
- Sigmundsson, Svavar (2012): *Íslensk samheitaorðabók* (3. ed.). Reykjavík: Forlagið.

Halldóra Jónsdóttir
 projektleder
 Árni Magnússon-instituttet for islandske studier
 Afdeling for leksikografi
 Laugavegur 13
 IS-101 Reykjavík
 halldo@hi.is

Þórdís Úlfarsdóttir
 hovedredaktør
 Árni Magnússon-instituttet for islandske studier
 Afdeling for leksikografi
 Laugavegur 13
 IS-101 Reykjavík
 disa@hi.is

En kognitiv semantisk analys av SAOB:s formler för betydelseförändring

Pär Nilsson

In my dissertation the method to describe semantic change used in the Swedish Academy dictionary (SAOB) is compared with cognitive semantic models of such a change. The work of SAOB is theoretically unbound in terms of overarching semantic theories, and the methods of describing word meaning and semantic change in the dictionary are rather developed for pragmatic reasons (such as consistency and economy etc.). The first volume of SAOB was published in 1898 and the latest in 2012. A lot has happened during this time within lexicographic theory as well as in general linguistics. To SAOB editors today the view of language and linguistics is to a large extent unlike the one the editors had in the late nineteenth century. Despite this fact the principles behind structuring and representing semantic information in the dictionary seem to have a lot in common with the theory based discussion in contemporary literature in the cognitive field. My starting point is to analyze the structure of articles in SAOB, in particular the use of specific labels used in the definitions, such as *bildl.* (figurative), *utvidgad anv.* (extended use) etc. My aim is to evaluate the pragmatic SAOB model and to investigate to what extent it can be interpreted to represent the way humans structure and categorize language and linguistic meaning. In this presentation I will demonstrate examples of how the SAOB method can be transferred to and interpreted through models derived from cognitive semantics.

1. Inledning och bakgrund

I mitt avhandlingsprojekt jämförs de metoder att beskriva betydelseförändring som kommer till uttryck i *Ordbok över svenska språket utgiven av Svenska Akademien* (SAOB) med kognitiva semantiska modeller för sådan förändring. Bakgrunden är avsaknaden av en övergripande explicit teori för den semantiska analys och beskrivning som genomförs på redaktionen idag.

Ordbokens form och struktur vittnar om de junggrammatiska ideal som rådde och var självklara och underförstådda vid tiden för ordbokens initiala skede. Teorin behövde inte sättas på pränt och diskuteras. Det historisk-filologiska synsättet var i det närmaste liktydigt med språkvetenskap vid den här tiden. Idag drygt 120 år senare när redaktörer ska ta sig an arbetet med ordboken och läsare ska tillgodogöra sig och acceptera innehållet i den är läget annorlunda och mycket har hänt inom den semantiska teorin.

Trots detta faktum verkar ordbokens principer för att strukturera och (re)presentera semantisk information ha mycket gemensamt med den teoribaserade diskussion som förs i den samtida litteraturen inom ramarna för den kognitiva semantiken. Att denna iakttagelse verkar rimlig bekräftas av bl. a. Geeraerts (2010) som redogör för hur man redan i den teoretiska diskussionen inom den historisk-filologiska skolan (som dominerade den semantiska teorin under andra halvan av 1800-talet och de första decennierna av 1900-talet) riktade uppmärksamhet mot psykologiska mekanismer bakom semantiska förändringar. Ett alltför ensidigt historiskt fokus, det diakrona studiets totala dominans, inom denna skola medförde så småningom att även dessa tankar om språket undanträngdes av andra synsätt som härefter kom att präglade den semantiska teorin.¹

Nu verkar det som att pendeln inom den akademiska semantiska teorin håller på att svänga, och den psykologiska aspekten är ofta åter i fokus. Intressant i sammanhanget är att det ofta inte är fråga om att de gamla teorierna och tankesätten har dammats av och börjat diskuteras igen, utan tankegångarna har i hög grad helt enkelt upptäckts på nytt (Geeraerts 2010:1). (Att det när det begav sig förra gången i huvudsak var de franska och tyska språken som gällde för texter inom den internationella (europeiska) semantiska teorin har sannolikt bidragit till att de äldre texterna har glömts bort och hamnat i bakgrunden).

2. Mål och metod

Hypotesen som drivs i projektet är att SAOB:s sätt att representera semantisk förändring, trots avsaknaden av en explicit teori, i mycket tangerar den

¹ Andra övergripande språkliga synsätt har till viss mån även påverkat SAOB-redaktionen och lämnat avtryck i ordboken, något som diskuteras i Nilsson (2012).

kognitiva semantikens ideal. Metoden är en analys av strukturen hos artiklar i SAOB, särskilt med fokus på användningen av de specifika formler eller fasta uttryck som används i definitionerna för att beskriva hur en (del)betydelse är relaterad till en annan, såsom *bildlig*, *utvidgad*, *överförd användning* osv. (i fortsättningen benämnda *definitionsformler*). Härigenom vill jag besvara frågor som: hur ser en (viss) betydelseförändring ut? Hur beskrivs den i SAOB? Är SAOB:s metod att beskriva förändringen teoretiskt relevant idag?

Nyckeln till analysen blir alltså ett studium av definitionsformler och redaktörernas etikettering av semantiska förändringsprocesser. I praktiken innebär detta sökningar i den digitala versionen av ordboken på strängar som ”bildl.”, ”utvidgad” osv. varefter närläsning sker för ett urval av antalet träffar. (En sökning på ”bildl.” ger drygt 25 000 träffar och ”utvidgad” drygt 9000, så begränsningar måste göras här.) I en första omgång består urvalet av 300 träffar. Detta urval fördelas jämnt på de fyra olika perioder i SAOB:s tillkomsthistoria som Larsson (2014) diskuterar och som i sig representerar olika omfångs- och ambitionsnivåer genom tiderna. Urvalet baseras på artiklarnas tryckår och tanken är att de utvalda artiklarna ska täcka och representera så mycket som möjligt av den aktuella perioden. För period 2 exempelvis (1920-1960²) har 75 träffar valts ut fördelade på olika decennier och i början mitten och slutet av denna period: de 25 första träffarna på bokstaven G (tryckår 1928), de 25 första på M (tryckår 1942) och slutligen de 25 första på P (tryckår 1952). Analysen sker i ljuset av modeller och teoretisk diskussion från litteratur på den kognitiva semantikens område och med målsättningen bl. a. att diskutera en teoretisk bas för en definitionsordbok av SAOB:s slag. Målet är att utvärdera den pragmatiska SAOB-modellen och undersöka i vilken utsträckning den kan sägas återspegla människans sätt att kategorisera och utveckla begrepp.

För att illustrera likheter och beröringspunkter mellan de i hög grad praktiskt orienterade ordboksartiklarna och diskussionen i den teoretiska litteraturen kommer jag i den här artikeln att ge några exempel på hur SAOB:s metod kan överföras till och tolkas genom modeller som härrör från den kognitiva semantiken. I det följande relateras ordbokstexten till olika aspekter av den kognitiva semantiska teorin, nämligen (i tur och ordning):

² Se Larsson (2014:65f) för närmare beskrivning av den aktuella perioden.

konceptuella metaforer, beskrivningen av det semantiska nätverket och avgränsningen av polysema ords betydelser, grammatikalisering såsom metaforisk överföring och, avslutningsvis, beskrivningen av metaforisk resp. bokstavlig förståelse av ett yttrande i tolkningsögonblicket.

3. Fyra nedslag i den teoretiska analysen i litteraturen. Från 1980-talet till idag – och SAOB:s beskrivningar

3.1 Lakoff & Johnson – konceptuella metaforer

Lakoff & Johnson (1980) lanserar tanken om att vardagsspråket är uppbyggt av konceptuella metaforer: “the essence of metaphor is understanding and experiencing one kind of thing in terms of another” (s. 5). Författarna ringar in olika typer av metaforer: strukturella, orienterande, ontologiska metaforer och talar om konceptuell metonymi. SAOB:s definitionsspråk och -formler analyseras i projektet bl.a. utifrån dessa begrepp. I Figur 1 nedan illustreras hur en semantisk förändring i SAOB kan analyseras med utgångspunkt i Lakoff & Johnsons olika metaforyper.

Strukturella metaforer är sådana som används systematiskt i vardagsspråket för att strukturera begrepp i termer av andra. Författarna ger bl. a. exemplet TIME IS MONEY. Vi inte bara talar om tid som om det var ett värdefullt ting, utan vi tänker och handlar också strukturellt i dessa termer, vilket följande uttryck i språket kan vittna om: *You’re wasting my time. How do you spend your time these days? He’s living on borrowed time.* (Lakoff & Johnson 1980:8). Hur ser det då ut i SAOB? Förklaras dessa strukturer i ordboken? Under uppslagsordet TID, sbst., moment 2 a beskrivs just denna syn på begreppet: tid som om det var en värdefull tillgång, och den specifika metaforen TID ÄR PENGAR anges som ett särskilt uttryck (med engelsk förlaga).

En särskild ställning bland metaforerna intar de orienteringsrelaterade. Dessa baseras direkt på våra erfarenheter av våra kroppar och det fysiska rummet. Konceptuella metaforer som HAPPY IS UP och SAD IS DOWN kan förklaras t.ex. genom att tillstånd av sjukdom o. dyl. ofta innebär sängliggande, medan vi som pigga och friska vanligen är uppe och står på benen. Som framgår av Figur 1 beskrivs dessa metaforer i SAOB under

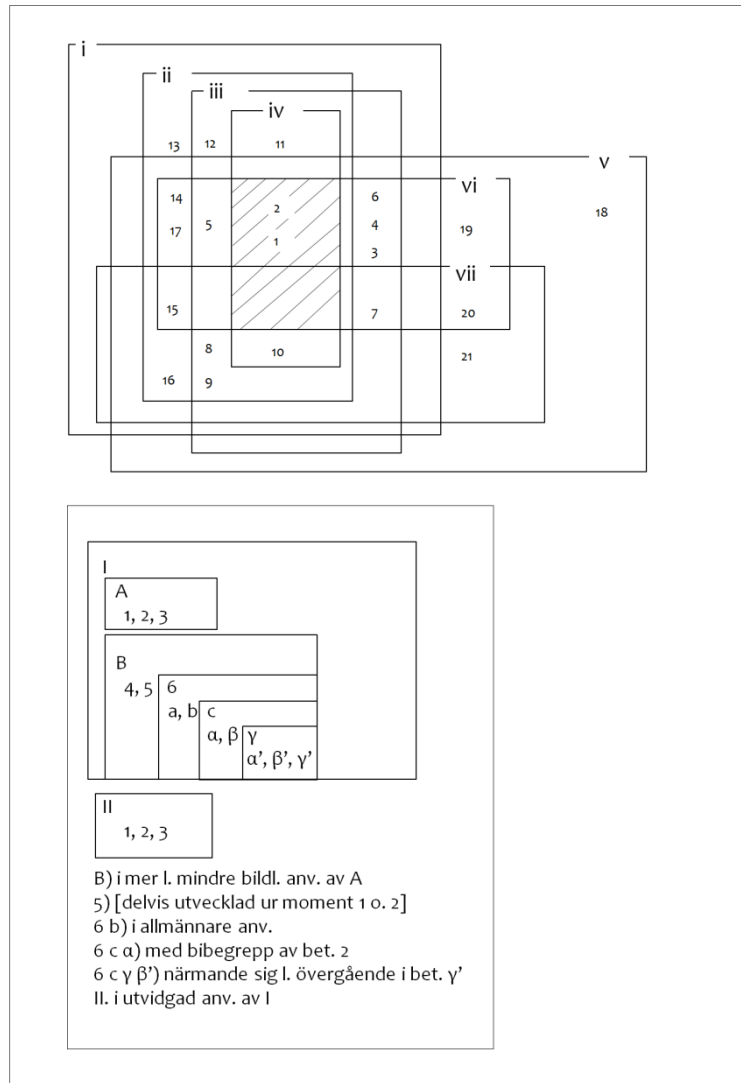
uppslagsorden UPPE och NERE (med hjälp av fasta definitionsformler) som oegentliga eller bildliga användningar av en befintlighet i ett visst läge.

Lakoff & Johnsons konceptuella metaforer:	Motsvarande beskrivning i SAOB:
* Strukturella metaforer: "time is money"	* TID: "om tiden ss. värdefull tillgång, särsk. i ordspr. l. ordspråksliknande uttr.; särsk. [efter eng. <i>time is money</i>] i uttr. <i>tid är pengar</i> "
* Orienteringsmetaforer: "happy is up, sad is down"	* UPPE: i oeg. l. mer l. mindre bildl. anv. ... angivande läge. [...] angivande att ngt l. ngn är i (full) verksamhet l. i gång o d; äv (i anv ss predikativt adj.): "på gott humör"
* Ontologiska metaforer: t.ex. <i>personification</i>	NERE: i bildl. l. oeg. bet. angivande befintlighet i visst läge l. tillstånd o. d. [...] [jfr eng. <i>down</i>] (ngt vard.) ss. predikativt adv., om person: i ett tillstånd av kroppslig o. (i sht) andlig depression; övergående i bet. av ett adj.: kroppsligt o. andligt försvagad, nedbruten, nedtryckt, nedstämd, deprimerad
* Metonymi	* Termer som inleds av "personi-" (personifikation, personifierad, personifiering osv.) används 479 ggr
	* Termer som inleds av "metonym-" används 336 ggr

Figur 1: Lakoff & Johnsons typer av konceptuella metaforer, med exempel (till vänster) jämförs med SAOB:s beskrivning av motsvarande process resp. motsvarande exempel (till höger).

De ontologiska metaforerna, som har att göra med vår tendens att tolka abstrakta och komplexa företeelser i termer av klart avgränsade och handfasta entiteter (Lakoff & Johnsonson 1980:25), företräds i Figur 1 av *personification* som ett särskilt slag härav. I SAOB omtalas ofta denna typ av bildspråk särskilt och i definitionerna kan den uttryckas på några olika sätt, bl. a. *personifikation*, *personifiering* eller *personifiera(d)t*. En fritextsökning i den digitala versionen av SAOB på frasen personi* ger 479 träffar som sedan är möjliga att närstudera. Nära bruket av metaforer står bruket av metonymier, vilket av Lakoff & Johnson definieras så här: "using one entity to refer to another that is related to it" (s. 35). Även metonymisk användning uppmärksammas i SAOB (under olika etiketter). En sökning på frasen metonym* ger 336 träffar i den digitala versionen av ordboken.

3.2. Geeraerts – det semantiska nätverket



Figur 2: Geeraerts analys och presentation av polysemin hos adjektivet *vers* (=färsk) utifrån beskrivningen i *Woordenboek der Nederlandsche Taal* (överst) jämförs med SAOB:s artikelstruktur och beskrivningar av semantiska relationer (underst).

Det lexikografiska perspektivet kommer till uttryck hos Geeraerts (1990). Författaren ifrågasätter rent linjär representation (1, 2, 3, osv.) av polysema ords betydelse(nyans)er i ordböcker och diskuterar alternativa lösningar. Den semantiska strukturen för det nederländska ordet *vers* (=färsk) analyseras mot bakgrund av beskrivningen i *Woordenboek der Nederlandsche Taal* (WNT) och en alternativ grafisk representation presenteras i en nätverksmodell. Geeraerts anser att den semantiska nätverksstrukturen går att utläsa ur WNT som den ser ut idag. SAOB:s hierarkiska struktur ihop med de hänvisningar och upplysningar om hur olika betydelser relateras, som också beskrivs i ordbokens spalter, motsvarar i mycket Geeraerts modell (Figur 2).

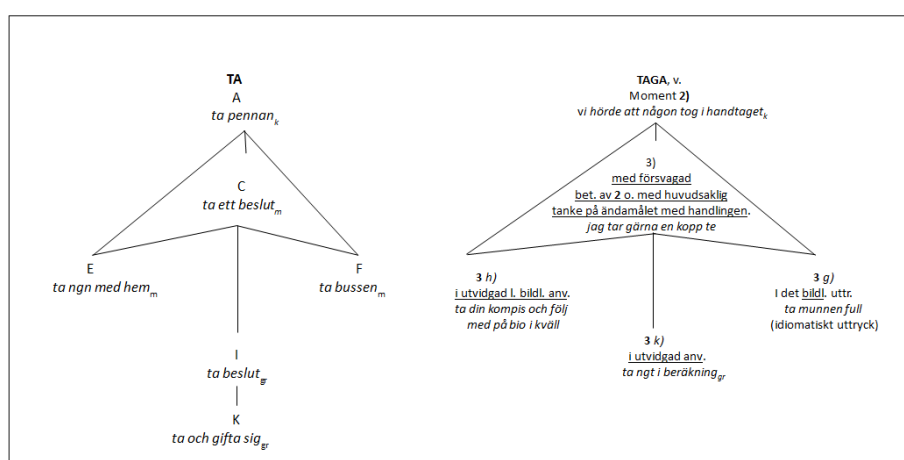
I Figur 2 återges Geeraerts modell över de olika betydelsemomenten schematiskt, utan sina respektive definitioner. Den typ av särskilda upplysningar om betydelsemomentens inbördes relation som förekommer i (samband med) SAOB:s definitioner och som återges i exemplet ovan ("i mer l. mindre bildl. anv.", "med bibegrepp av bet. 2" osv.) bryter upp det strikta hierarkiska systemet och indikerar en mer korsande och överlappande struktur. De båda presentationssystemen är därmed jämförbara.

3.3 Ekberg – verbet *ta(ga)*

Lena Ekberg (1993) beskriver svenska ord och betydelser. Ekberg analyserar metaforiska och grammatikaliserade användningar av verbet *ta* och hävdar att de båda är resultatet av en och samma kognitiva process, nämligen metaforisk överföring från en domän till en annan. Ekberg presenterar olika användningar av verbet i föreställningsscheman. SAOB:s beskrivning av TAGA, v. (med särskilt fokus på de etablerade definitionsformler som används) kan överföras till ett motsvarande schema. Ekbergs analys och diskussion indikerar det diakrona studiets förtjänster för förklaring och förståelse också vid synkron analys av begreppsbildningen.

I SAOB:s definition av TAGA moment 3, används just de etablerade definitionsformlerna i hög grad för att illustrera inbördes relationer mellan delbetydelser. Om man accepterar att en användning som Ekbergs *ta beslut* kan motsvaras i fråga om grammatikalisering av SAOB:s *ta ngt i beräkning* (TAGA moment 3 k; härunder redovisas bl.a. även sådana uttryck som *ta i betänkande*, *ta i beskydd*, *ta i betraktande*), och att *ta ngn med hem*

motsvaras av *ta din kompis och följ med oss på bio ikväll*, blir en analys av definitionsformlerna intressant ur diakront perspektiv. I det förra fallet beskrivs användningen i SAOB som utvidgad eller bildlig och i det senare som (enbart) utvidgad. Genom att undersöka definitionsformler (och andra medel som står till buds för att förklara kopplingen mellan betydelsemoment i ordboken) går det att bedöma om grammatikaliserade användningar av ett ord systematiskt beskrivs annorlunda än metaforiska användningar.

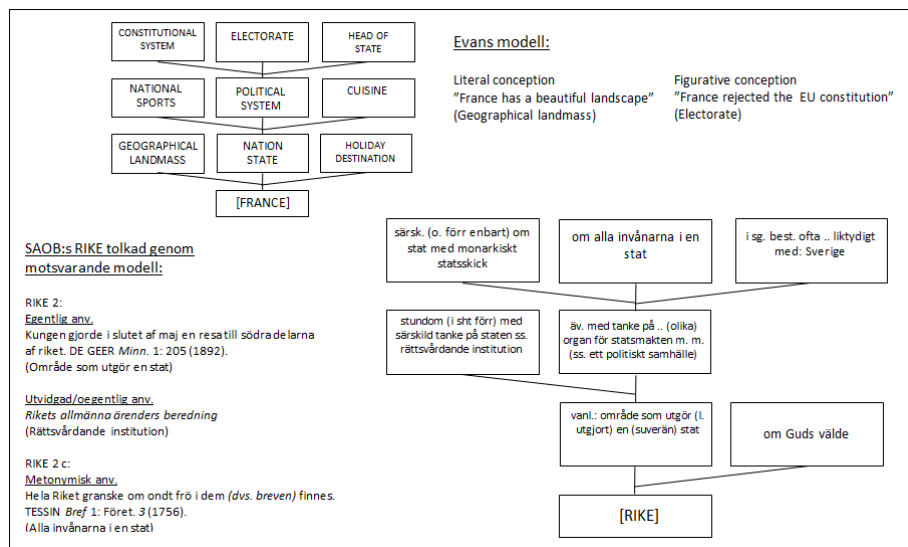


Figur 3: Ekbergs föreställningsschema över konkreta, metaforiska och grammatikaliserade användningar av verbet *ta* (till vänster) jämförs med delar av SAOB:s beskrivning av TAGA,v överfört till motsvarande schema. k= konkret, m= metaforisk, gr= grammatikaliserad.

3.4 Evans – tolkningsprocessen

Ett exempel ifrån den språkliga insidan utgörs av Evans (2009). Evans försöker inte enbart återge strukturen eller principerna för de semantiska relationerna hos polysema lexem med utgångspunkt i de spår i språket som ordanvändningen lämnar efter sig, utan vill även förklara hur själva den mentala processen ser ut vid ögonblicket då tolkningen och förståelsen av ett yttrande äger rum. Bl.a. diskuteras hur begreppsbildning skiljer sig åt i fråga om bokstavlig och figurativ förståelse. I en förenklad modell illustreras hur man utgår ifrån primära, konkreta begrepp i möjligaste mån innan man tar en figurativ omväg ("uppåt") till förståelse (2009:288f). SAOB:s hierarkiska

struktur och (logisk-)kronologiska presentationssystem återger denna aspekt, att bokstavlig tolkning föregår figurativ.



Figur 4: SAOB:s grafiska representation av utvidgad, oegentlig och metonymisk användning (nedan) tolkad genom Evans illustration av "figurative conception" (ovan).

Så länge en användning eller betydelse uppfattas som figurativ sker tolkningen *via* den bokstavliga betydelsen; i annat fall utgör användningen inte längre en delbetydelse utan har utvecklats till en självständig och då uppfattad som (ny) bokstavlig betydelse. SAOB:s perspektiv är nu inte den individuella språkbrukarens mentala språk, utan analysen gäller det kollektiva språkbruket och resultatet av språkbrukarnas olika användningar (även om analysen sker i huvudet på redaktören som utför den). Ändå kan det vara förtjänstfullt att understryka att den historiska utvecklingen hos ordet eller betydelsen, som representeras av den grafiska återgivningen i ordboken, också har beröringspunkter med den individuella, synkrona tolkningsprocessen.

4. Slutord

Det torde vara önskvärt att knyta framtida ordboksarbete till en tydlig och uttalad semantisk teori för att förankra det vetenskapliga arbetet, för att systematisera det och för att spara tid. Med stora penseldrag har jag här försökt jämföra den lexikografiska praktiken (med rötterna i en språksyn från slutet av 1800-talet) med ett teoretiskt ramverk från modernare tid. Jag har bara skrapat på ytan i ett försök att visa tydliga beröringspunkter mellan praktik och teori. I vilken utsträckning som SAOB:s beskrivningar av semantiska förändringsprocesser mera i detalj tangerar kognitiva semantiska ideal och hur pass konsekvent man har varit i sitt utförande genom åren återstår att ta reda på.

Litteratur

Ordböcker

SAOB (1898-) = *Ordbok över svenska språket utgiven av Svenska Akademien*. 1-. Lund.

Annan litteratur

Ekberg, Lena (1993): Verbet *ta* i metaforisk och grammatikaliserad användning. I: *Språk och Stil. Tidskrift för svensk språkforskning* NF 3, 105-139.

Evans, Vyvyan, (2009): *How words mean: lexical concepts, cognitive models, and meaning construction*. Oxford, Oxford University Press.

Geeraerts, Dirk (1990): The lexicographical treatment of prototypical polysemy. I: Tsohatzidis, Savas. L. (red.): *Meanings and Prototypes. Studies in linguistic categorization*. London: Routledge, 195-210.

Geeraerts, Dirk (2010): *Theories of lexical semantics*. Oxford: Oxford University Press.

Lakoff, George & Mark Johnson (1980): *Metaphors We Live By*. Chicago & London: The University of Chicago Press.

- Larsson, Lennart (2014): En ”mer l. mindre stor” stor ordbok – om variationerna i SAOB:s omfång och ambitionsnivå. I: *LexicoNordica* 21, 61-80.
- Nilsson, Pär (2012): Språkteoriens avtryck i SAOB. I: Eaker, Birgit, Lennart Larsson & Anki Mattisson (red.): *Nordiska studier i lexikografi* 11. *Rapport från Konferensen om lexikografi i Norden*, 479-487.

Pär Nilsson

Licentiat i Nordiska språk
Lunds universitet
Helgonabacken 12
SE-221 00 Lund

Ordboksredaktör
Svenska Akademiens ordboksredaktion
Dalbyvägen 3
SE-224 60 Lund

Par.Nilsson@svenskaakademien.se

NAOB: På vei mot en nasjonalordbok for bokmålet: Nyordprosjektet ”2014 nyord i 2014”

Carina Nilstun

This article reports from the campaign ”2014 neologisms in 2014”, conducted by the editors of The Norwegian Academy’s Dictionary (NAOB), as a step towards publishing a national dictionary for the Norwegian variety bokmål. The Academy’s dictionary will be published on <naob.no> in 2017. The campaign aimed to involve our future users, inviting them to search for neologisms with us. In a weekly feature in a popular radio program throughout the whole 2014, a winner was announced, and by the end of the year, our goal was reached: 2014 neologisms were collected, lexicographically prepared and published.

1. Bakgrunn

I 2014 feiret Norge 200 årsjubileum for Grunnloven. Det var planlagt at to nasjonale ordboksverk skulle bli ferdige til jubileet: *Norsk ordbok 2014*, som beskriver det norske folkemålet og det nynorske skriftmålet, og *NAOB*, som beskriver bokmål/riksmål.

Finansieringen strakk ikke til for bokmålets del, og publisering ble skjøvet til 2017. Fra og med 2015 har prosjektet imidlertid fått økt sin støtte fra staten betraktelig, og staben er nå noe mer enn doblet. Da forsinkelsen var et faktum, var det allikevel et ønske om å gjøre noe i jubileumsåret, nemlig å gi en forsmak på den digitale ordboken. Men hvordan? Et alfabetutsnitt? Et utvalg fra hver ordklasse? Slike løsninger ville vært enveiskommunikasjon, og en gjest på nettsiden ville ikke ha noen grunn til å komme tilbake før i 2017. Vi ville ha noe som kunne engasjere og være dynamisk – uten at det ble for mye av et sidespor for arbeidet frem mot 2017. Valget falt på nyord. Nyord engasjerer, og det er et område i vårt materiale som trengte et reelt løft.

Vårt prosjekt går ut på å revidere en diakron ordbok (Norsk Riksmålsordbok), og i første omgang er vårt format elektronisk. For en diakron papirordbok vil tidsrommet man beskriver nødtvungent være avgrenset

i begge ender. Våre rammer lar oss ha den bevegelige størrelsen ”dags dato” som endepunkt. Arbeidet med å fange opp neologismer er således viktig, og ved å engasjere publikum i dette fikk vi samtidig effektivisert sankingen. Vi lot oss inspirere av jubileumsåret da vi skulle sette oss et mål for antall nyord vi skulle samle inn og presentere, og dermed var prosjektets tittel klar: *2014 nyord i 2014*. Med nyord mener vi ikke bare nye *ord*, men også nye betydninger og nye uttrykk.

Kampanjen ble ført an av Helene Uri, kjent og respektert lingvist, forfatter og styremedlem i NAOB. Hver torsdag gjennom året, i en fast spalte med egen kjenning i NRK Radios populære formiddagssending Nitimen, snakket Helene om nyord og inviterte lytterne til å sende inn sine nyordsforslag. I hver sending ble et knippe finalister og én vinner fra forrige ukes bidrag presentert. Vinnerne ble hedret ved navns nevning, både på radio og på nettstedet. Samlingen av nyord vokste fra uke til uke, og vi passerte 2014 nyord oppunder jul. Per høsten 2015 er antallet rundt 2200.

2. Hva er et nyord?

En del av forslagene som ble sendt inn var intuitivt gode nyord. Noen kjente vi allerede, og noen hadde vi allerede registrert og bearbeidet i redaksjonen. Andre var helt nye for oss, men ble bekreftet som gode nyord ved søk i relevante tekstbaser.

Et eksempel på et godt nyord er *ureist*, særlig brukt i forbindelsen *ureist mat*. Utgangspunktet er *kortreist mat* (mat med kort vei fra produsent til forbruker). *Langreist mat* har oppstått som motsetning til *kortreist mat*, og til sist kommer ordet for mat som ikke har reist i det hele tatt, mat som er produsert (gjerne dyrket) akkurat der den skal spises (i kjøkkenhagen, i bakgården) – *ureist mat*. Dette er en morsom og kreativ orddannelse, og alle tre (*ureist*, *kortreist* og *langreist*) er kommet med i NAOB.

Vår involvering av publikum brakte også mange ord som definitivt ikke kunne passere. Hovedmengden av disse var private dannelser – enten ord som bare brukes blant en svært eng gruppe mennesker, som en familie eller en vennegjeng, eller rene ideer, uten noen som helst bruk å vise til. Noen oppfattet altså nyord rent bokstavelig – at ordene skulle være ubrukte, og at NAOBs nyordprosjekt dermed var en arena for å lansere egne ord. Noen

ønsket en bekreftelse fra oss på at ordet de hadde funnet på, burde bli en del av språket, noe som er en uvant tilnærming for oss som jobber med språk. Noen eksempler, med innsenders forklaring i parentes: *Julegavebadekar* (et babybadekar du oppbevarer opp-pakkede julegaver i), *kontrollfjerner* (da barna var små sa de kontrollfjerner til *fjernkontrollen*), *løke* (når man legger noe i flere lag. ”Skal jeg finne flere steder å henge opp håndklærne, eller skal jeg bare løke dem over stolen?”), *mammapedia* (når mine voksne barn ringer for å spørre meg om ting (mamma kan det meste hos oss)), *varmiator* (bedre enn *radiator* som ikke gir noen hint om hva ordet beskriver).

En annen gruppe ord vi unngår, er ord man kjenner som nyord, men som ved nærmere undersøkelser overveiende gir metabelegg. Dette er nærmere tematisert hos Jansson (2016). Et eksempel er *banansvenske*, som oppstod i 2011 med utgangspunkt i svensk ungdom som kommer til Norge og jobber med å skrelle bananer på fabrikkens Banos. Beleggene er av typen ”nyordet *banansvenske* har ambassadøren fått med seg, men hun liker det ikke” og ”en T-skjorte med nyordet *banansvenske* på brystet”.

2.1. Kriterier

Noen ord er altså enkle å rubrisere som enten nyord eller ikke nyord, men i de fleste tilfeller trenger man noen faste kriterier å gå etter. Et ord som brukes for første gang, anses ikke som et nyord. Der og da er bruk av et ord som ikke tidligere har vært brukt, en adhocdannelse, eventuelt et adhoclån, og dets videre status avhenger av i hvilken grad bruken gjentas.

Tid er naturligvis viktig når man skal definere hva som er et nyord. Vi har ikke satt noen skarp grense, men opererer grovt sett innenfor de siste 15–20 år. Også rom/domene er relevant. Vi snakker om nyord i *allmennspråket*. Ord som i lengre tid har vært brukt i bestemte miljøer, især fagmiljøer, som så går over til allmennspråket, kan hos oss regnes som et nyord. Vi er da over på *frekvens*, et annet viktig kriterium. Det skal en viss bruk til, før noe kan sies å være et nyord. Frekvens skiller nyordene fra adhocordene. I en pragmatisk vurdering med disse aksene til veiledning foregår vurderingen av hva som gjelder som nyord hos oss.

Tid og frekvens står som hovedkriterier når man snakker om nyord, og mange avgrenser nok tiden strengere enn det vi har gjort i vårt prosjekt. Vårt *perspektiv* har ikke vært verken ordene i seg selv eller tiden, slik det for ek-

sempel er når Språkrådet kårer årets nyord¹. Hos oss har leksikografi-perspektivet på det nye vært overordnet: Vi har jaktet ord av nyere dato som ikke allerede er leksikografisk beskrevet, og omsatt dem til leksikografi. Ord som allerede er beskrevet i ordbøkene *Bokmålsordboka* (2005), *Norsk ordbok med 1000 illustrasjoner* (2008) eller *Norsk Ordbok* (1966–2015), gjelder ikke som nyord hos oss. Dette ga også mening for delpubliseringen: Både form og innhold er nytt – det man finner på <naob.no> frem mot lanseringen i 2017, skal man ikke finne andre steder.

En del av ordene i nyordsamlingen oppfattes ikke som nye. Dette forklares blant annet med at vi baserer oss på det skriftlige. Noen ord har tidligere levd bare i det muntlige, i periferien av allmennspråket eller med mer spredt og tilfeldig bruk, for eksempel *jentelus/guttelus*, *innestemme/utestemme*. Dette er ord som vi alle kjenner, og som har levd lenge i det muntlige. De er sparsommelig belagt i skrift, og har ikke funnet veien inn i noen ordbok før nå. Ordtypen nevnt over finner man helst i skjønnlitteratur som beskriver barns verden.

jentelus substantiv

BØYNING en

BETYDNING OG BRUK

SPØKEFULLT fiktiv lus hos jenter som gjør at gutter holder seg unna dem, særlig virksom i barneskolealder | jf. [guttelus](#)

SITAT

- *hun stilte seg omtrent oppå meg. Jeg kunne nesten kjenne jentelusa hoppe over og lande på T-trøya mi* (Marit Nicolaysen *Svein og rotta i Syden* 48)

Publikasjon: [Det Norske Akademis Store Ordbok](#)

[Send en kommentar til denne artikkelen.](#)

Ordene *brytterygg* og *dynegrøt* kan nok også være en god del eldre enn det de skriftlige beleggene tilsier.

Ordet *bygulv* har opprinnelig vært brukt innen domenet arkitektur/byplanlegging, med samme billedbruk som *byrom*. Første belegg er i en kronikk forfattet av en landskapsarkitekt i 1990. Gradvis har bruken økt, og ordet har beveget seg over i allmennspråket. Kriteriene tid, frekvens og do-

¹ <www.sprakradet.no/Vi-og-vart/Prisar_og_karingar/Arets-ord/>.

mene (allmennspråk) er da tilfredsstilt. Siste kontrollpost blir om ordet tidligere er beskrevet i en allmennordbok.

bygulv substantiv

BØYNING et

BETYDNING OG BRUK

[dekke](#) som kler bakken i [byrommet](#)

SITATER

- *broleggingen med Iddefjordgranitt i første halvdel av dette århundret skapte utsøkte bygulv (Aftenposten Aften 03.01.1994/18)*
- *gamle og slitne asfalttepper erstattes med presis og renskåren granitt som skikkelig bygulv (Morgenbladet 22.04.2005/20)*

Publikasjon: [Det Norske Akademis Store Ordbok](#)

[Send en kommentar til denne artikkelen.](#)

2.2. Konkurransévinnere

Hva ble så ansett som verdige vinnere? Dette var en subjektiv vurdering, fristilt fra kriterier om for eksempel frekvens. Dels fulgte vi årets begivenheter. Verbet *palle* og substantivet *skrell* kom i forbindelse med OL i Sotsji, mens *dekkhotell* var vinnerordet når tiden var inne for å skifte fra sommerdekk til vinterdekk på bilen. Vi forsøkte også å ha god variasjon – ulike ordklasser, ulike orddanningsmåter. Gode, morsomme, positive, nyttige ord ble løftet frem – det skulle jo også lages god radio av dette.

3. Presentasjon

Nyordbasen er avgrenset, og en bruker vet ikke hva som befinner seg der. Bruken vil være av mer utforskende art enn bruken av en allmennordbok, hvor man oppsøker verket med et konkret ord i tankene. Et søkefelt alene er derfor ikke fyllestgjørende som atkomst til dette stoffet, en alfabetisk liste må til for å gi uttømmende tilgang. I tillegg er nyordstoffet godt egnet til å grupperes ut fra en rekke kryssende trekk. Nettstedet byr derfor på siden *Utforsk nyordene*. Det er satt opp en rekke akser, og hvert nyord kan ak-

sesseres ad flere veier, for eksempel orddanningsmåte, stil og fagområde. Vi skal se nærmere på et utvalg av disse aksene.

3.1. Utforsk nyordene

Kategorien Dagligliv har mange underkategorier, som adferd, klær, skole, barn (barns liv), lek og spill, innredning, mat, språk og lesing, utdanning og utstyr. Her følger noen eksempler fra et utvalg av disse:

Adferd: *alfahunn, ammepoliti, avstikk, curlingforeldre, dobbeltkommunisere*

Klær: *baggy, buff, kompesjonstights, manneveske, mesh, shoppestopp, skalljakke*

Mat: *cupcake, dobbeltdippe, feriemage, fredagstaco, kjøkkenøy, ortoreksi, pizzastein, pod, signaturrett, slow-food, sous vide, veggis*

Utstyr: *bålpanne, bæresjal, gorillaglass, gruski, hentesett, joggevogn, pulsbelte, tarp*

Assosiasjonene er forholdsvis vide, slik at *feriemage* sorterer under mat, og *shoppestopp* under klær. Disse kategoriene knyttes til det enkelte ord i tagger som ikke vises. Hvert ord kan ha mange tagger, både fra kryssende kategorier og innen samme kategori. *Shoppestopp* har således tre fag/domene-tagger, *shopping, klær, privatøkonomi*.

Fra *Utforsk nyordene* kan man dykke ned i andre inndelinger, for eksempel ordets alder sett i forhold til referansens alder. Mange nyord kommer som følge av teknologisk eller samfunnsmessig utvikling. Ikke bare ordet, men også fenomenet er da nytt (*nettbrett, fedrekvote*). I andre tilfeller er fenomenet eller gjenstanden allerede benevnt, det nye ordet tar over for det gamle, eller brukes parallelt med det, kanskje med en annen valør (f.eks. *miljøbod* for *søppelbod*, *spisse* (*organisasjonen*) for *nedbemannen*). Den tredje gruppen her er fenomener som er gamle og velkjente, men som man tidligere ikke har hatt eget ord for. Et eksempel er *monobryn*, som vi tidligere beskrev ikke-leksikalisert med to ord, sammenvokste øyenbryn.

I kategorien *Ordklasse* er det som forventet stor overvekt av substantiver, faktisk nesten 90 % av det totale antall. Tre andre ordklasser skal vi se litt nærmere på: suffiks, sammensetningsledd og konjunksjon².

Vi har ett suffiks blant nyordene, *-ish*, lånt fra engelsk. Bruken er velkjent – især muntlig, men det er også godt med skriftlige belegg. Dette suffikset er svært produktivt, og man ser og hører det i stadig nye avledninger. Det har to betydninger, den første angir at noe ligner eller minner om noe annet, synonymt med *-aktig*, den andre modifierer et tall (tid eller alder), og har betydningen ‘omtrent, cirka’. Ordbokartikkelen har sitater med *Edit Piaf-ish*, *professor-ish* og *tegneserie-ish* for betydning 1 og *han var 20-ish* for betydning 2. Siden vi dokumenterer snarere enn normerer, er en anseelig mengde av nyordene engelske lån.

Sammensetningsledd er egen ordklasse i NAOB. Blant nyordene har vi ni sammensetningsledd. Et av dem er *monster*, brukt som førsteledd med betydning '(svært) stor; gigantisk', for eksempel *monstermast*, *monsterhit*.

Vi har også to nye konjunksjoner. De er (ikke overraskende) ikke nye ord, men nye betydninger. De to er slash og skråstrek, og brukes som følger:

en skrekk slash komediefilm (Fanaposten 31.01.2014/14)

traileren til "Rosemary's Baby" ble lansert på nettet i forrige uke, så seerne kan begynne å glede skråstrek grue seg i god tid før TV-premierer (Dagsavisen 22.04.2014/23)

Den siste kategorien i *Utforsk nyordene* som skal nevnes her, er *Forfatter-skap*. Nyordene skal altså ledesages av skriftlig belegg. Flesteparten av beleggene er fra aviser og tidsskrifter, men noe er fra skjønnlitteratur og sakprosa. Mot slutten av 2014 hadde sitatmengden fra skjønnlitteratur og sakprosa blitt såpass omfattende, at tiden var inne til å opprette denne kategorien. Her finner man blant andre kjente norske forfattere som Karl Ove Knausgård, Tove Nilsen, Vigdis Hjorth, Jo Nesbø og tegneserieskaperen Frode Øverli. Koblingen til vårt grunnlag, Norsk Riksmålsordbok, er da tydelig. Riksmålsord-

² Suffiks og sammensetningsledd inngår ikke som ordklasser i tradisjonelle ordklasseinndelinger. I leksikografiens praktiske verden hvor ordartiklene så langt det lar seg gjøre, skal gis ordklassetilhørighet, forekommer det at man går utover den tradisjonelle ordklassmengden. I NAOB er sammensetningsledd, suffiks (og prefiks) ordklasser.

boken er en litterær ordbok – i den betydning at den baserer seg på det skriftlige (til forskjell fra det muntlige), og forbindes nok særlig med skjønnlitteraturen.

4. Veien videre

I løpet 2014-kampanjen dannet det seg en gruppe ivrige bidragsytere som sendte inn regelmessig og bidro svært godt i innsamlingen. Vi annonserte derfor underveis at vi ved årets slutt ville trekke frem de fem skarpeste nyordjegerne – det vil si de fem som hadde fått flest av sine forslag med i ordboken. Det ble fire kvinner og én mann, og vi inviterte dem til en markering av kampanjens avslutning, hvor vi også var så heldige å få med oss kulturminister Thorhild Widvey. Disse ivrige nyordjegerne og flere med dem fortsetter å sende inn sine forslag også i 2015, slik at det fortsatt publiseres nyord på <naob.no>, og fortsatt kåres vinnerord med ujevne mellomrom frem mot den fulle publiseringen. De seneste vinnerordene er *snikskryte*, *pappakropp*, *taleleif/trykkeleif* og *vennebenk*.

Som en ytterligere smakebit på hva som kommer i 2017, rommer nettsiden nå også spalten Ord fra kjelleren. Med utgangspunkt i gamle norske spenningsføljetonger, nærmere bestemt dem som ble publisert på avissidenes ”kjeller”, relanserer vi ord som har forsvunnet fra dagligtalen. Ukens ord presenteres med tilhørende NAOB-artikkel, relevant språklig og kulturell kontekst, samt en faksimile av føljetongepisoden som det aktuelle ordet er hentet fra. Vi aktiviserer brukerne ved å oppfordre dem til å bidra med sitater fra Nasjonalbibliotekets enorme digitale arkiv. Dette viser da den andre enden av NAOB, de gamle ordene, ordene som er havnet ”i kjelleren”.

NAOB er nå å finne på Facebook og Twitter, hvor nye kjellerord og nye nyord annonseres. Dere er velkommen til å følge oss på sosiale medier og til å besøke <naob.no>. Nyordforslag sendes inn via skjemaet på nettstedet.

5. Avsluttende ord

Prosjektet *2014 nyord i 2014* har krevd en god del ressurser i redaksjonen, samtidig ser vi mange gevinster:

- Vi har måttet jobbe med selve publiseringen, og fått testet ut publiseringsplattformen.

- Vi har gitt ordboken et viktig og nødvendig løft i den nyere enden av tidsrommet den dokumenterer.
- Vi har forhåpentlig fått til en viss profilendring: fra den ”tilbakeskuende” Riksmålsordboken til en ordbok som er dynamisk og moderne, både i format og innhold.
- NAOB har fått publisitet og blitt mer synlig.
- Vi har fått erfaring i å ha kontakt med publikum. Vi tror interaksjon mellom redaksjon og brukere blir viktig for ordbøker i fremtiden.

Litteratur

Ordbøker

Bokmålsordboka (2005): Boye Wangensteen (red.). 3. utgave. Oslo: Kunnskapsforlaget.

Norsk ordbok med 1000 illustrasjoner (2008): Tor Guttu (red.). 2. utgave. Oslo: Kunnskapsforlaget.

Norsk Ordbok. Ordbok over det norske folkemålet og det nynorske skriftmålet 1-12 (1966–2015). Oslo: Det Norske Samlaget.

Annen litteratur

Jansson, Håkan (2016): *Från ”tandborstord” til ”memilord” – om nyord och deras belägg*. I: Asgerd Gudiksen & Henrik Hovmark (red.): *Nordiske Studier i Leksikografi 13. Rapport fra 13. Konference om Leksikografi i Norden København 19.–22. maj 2015*. København: Nordisk Forening for Leksikografi, dette bind.

Carina Nilstun
leksikograf, MA
Kunnskapsforlaget
Gullhaug torg 1
Postboks 4432 Nydalen
NO-0403 Oslo
carina.nilstun@kunnskapsforlaget.no

Fra begrebsordbog til sprogteknologisk ressource: verber, semantiske roller og rammer – et pilotstudie

Sanni Nimb & Bolette Sandford Pedersen

This paper describes a method of compiling a lexicon of Danish semantic frames within the model of the Berkeley FrameNet (BFN). Large groups of near-synonymous verbs and verbal nouns, including multiword units, within the domains of communication and cognition are identified and extracted from the source manuscript of a newly published Danish thesaurus. Each word or expression is then assigned an appropriate frame from BFN. The fact that words within the same domain all belong to a manageable subset of frames in BFN makes it possible to map a high number of words to their corresponding frames simultaneously. In a forthcoming annotation project where words within the same two domains are already identified in the corpus, the idea is to pre-annotate with the frames in our lexicon, leaving afterwards human annotators to confirm the frame and test whether it is possible to identify the BFN semantic roles described for English in the Danish text. Our method reveals some interesting divergences between the semantic divisions established in the thesaurus in contrast to the ones found in BFN, showing that the two resources contribute with different types of linguistic information and thereby constitute a useful supplement to one another.

1. Baggrunden for pilotstudiet

I forskningsprojektet *Semantic Processing across Domains*, et 3-årigt samarbejdsprojekt mellem Københavns Universitet og Det Danske Sprog- og Litteraturselskab, støttet af Det Frie Forskningsråd, eksperimenteres der med forskellige former for semantisk opmærkning af danske, genre-mæssigt meget forskellige tekster, som grundlag for en efterfølgende udvikling af automatiske processeringsmetoder til brug for forskellige former for sproglig betydningsanalyse. Der skal bl.a. eksperimenteres med opmærkning af semantiske rammer inden for den i princippet sproguafhængige model der er etableret i det amerikanske FrameNet-projekt *Berkeley FrameNet* (BFN), og

som allerede er bevist velegnet til overførsel til andre sprog, fx spansk, japansk, svensk og fransk, jf. Boas (2009). Fordelen ved at benytte en standard som BFN til opmærkning er at man dermed får adgang til at overføre de sprogteknologiske resultater indenfor automatisk betydningsanalyse der er opnået med anvendelsen af BFN på andre sprog. Denne artikel beskriver vores arbejde i forskningsprojektet med at etablere et BFN-rammeleksikon for dansk til brug for det planlagte opmærkningsarbejde. Vi beskriver hvordan rammeleksikonnet opbygges på basis af den semantiske gruppering af det danske ordforråd som man finder i Den Danske Begrebsordbogs grundmanuskript (DDB 2014). Efterfølgende skal leksikonnet bruges til præannotering af vores korpustekster således at annotøren kan nøjes med at godkende den automatisk tildelte ramme, evt. blandt flere mulige forslag, og derefter fokusere på at opmærke rammens tilknyttede semantiske roller sådan som de udmøntes i den danske tekst. Det særlige ved vores metode er at de mulige, semantiske roller som annotøren skal identificere i teksten, overtages direkte fra BFN's rammebeskrivelse. Vi har nemlig en hypotese om at det rolleinventar som er fundet i engelske tekster og derefter omhyggeligt beskrevet i BFN, sandsynligvis også vil dække de roller der er relevante at opmærke i danske tekster, og at man derfor kan spare den tidskrævende arbejdsgang det ville være at beskrive rolleinventaret for hvert verbum i selve det danske leksikon inden annoteringen påbegyndes.

1.1. Forskelle og ligheder mellem et FrameNet og en begrebsordbog

Både en begrebsordbog som DDB og et formelt semantisk leksikon som BFN beskriver handlinger og situationer. Et FrameNet udgør en formaliseret delmængde af en begrebsordbog, blot uden en overordnet emneinddeling. En såkaldt ramme (frame) i BFN, inklusive dens tilknyttede leksikalske enheder, svarer til den række af ord i et afsnit i en begrebsordbog der beskriver handlingerne eller hændelserne inden for det emne afsnittet omhandler. Omvendt mangler en begrebsordbog en præcis betegnelse for netop denne gruppering af ord – afsnittets navn er typisk bredere formuleret og dækker også andre former for begreber i afsnittet, ikke kun handlinger og hændelser. Et FrameNet indeholder desuden en formaliseret beskrivelse af disse handlinger og hændelser. For det første navngives hver enkelt handlings- og hændelsesgruppe med en præcis og entydig rammebetegnelse,

og denne suppleres yderligere med oplysninger om de elementer der kan indgå i den i form af en række semantiske roller fra et afgrænset inventar, fx en 'Speaker', en 'Message' og en 'Addressee'. Derudover listes der ved hver ramme de leksikalske enheder der kan have den pågældende rammebetydning (svarende til rækken af ord og udtryk i gruppen i DDB). Et eksempel er BFN-rammen 'Statement' der har den leksikalske enhed *talk* tilknyttet. De leksikalske enheder kan selvfølgelig være polyseme inden for BFN, dvs. have flere rammebetydninger, fx har den engelske verbum *talk* udover 'Statement' også fire andre rammer, fx 'Discussion' (svarende til at verbet *tale* står i mange forskellige afsnit i DDB). Men der er også tilfælde hvor en leksikalsk enhed kun har én rammemulighed, således har *think up* kun rammen 'Coming_up_with'.

I BFN-projektet (der fortsat udbygges med rammer) fokuseres der primært på den detaljerede opbygning af rammer og roller for det engeske sprog, og det har derfor en forholdsvis lav leksikalsk dækningsgrad (13.099 ord), men til gengæld 1073 rammer med tilhørende formaliserede rollebeskrivelser. I DDB fokuseres der lige modsat på selve det danske ordforråd. Vores metode går ud på at kombinere de to ressourcer i opbygningen af et dansk FrameNet, så vi opnår det bedste af to verdener. Vi udnytter det store inventar af allerede beskrevne rammer fra BFN, men lader ordforrådet i DDB udgøre de leksikalske enheder, idet vi i øvrigt overfører alle roller og relationer som er beskrevet i rammerne, direkte fra BFN. Vi tager udgangspunkt i DDB's tematiske inddelinger og rækker af nærsynonyme ord og udtryk idet vi udnytter at begrebsordbogens grundmanuskript indeholder formaliserede oplysninger om ordenes semantiske kategori (dvs. om de fx er handlinger eller hændelser). I det svenske FrameNet-projekt (se Friberg & Dannells 2015) benytter man også et semantisk leksikon som udgangspunkt, nemlig SALDO; herfra linkes der på samme måde til BFN. Men SALDO er organiseret efter indbyrdes semantiske relationer mellem ordene, ikke efter temaer; på den måde adskiller det sig en del fra en begrebsordbog, jf. Borin & Forsberg (2009). Bicks udvikling af en dansk framenet-lignende ressource (se Bick 2011) bygger heller ikke på et tematisk udgangspunkt, men på et constraint-grammar-baseret valens-leksikon suppleret med information fra det danske WordNet DanNet. I ressourcen defineres et sæt af i alt 38 semantiske roller som ikke er direkte kompatible med BFN's roller.

2. Udvalgelse af ordforråd til pilotstudiet

Afgrænsningen af ordforrådet til vores pilotleksikon bygger på de resultater vi har opnået i den første del af forskningsprojektet, nemlig den del der omhandler leksikalsk korpusopmærkning med betydningskategorier. Udover en eksperimentel opmærkning med et lille udvalg af Den Danske Ordbogs (DDO) og det danske WordNet DanNets finkornede betydningsinventar (se Pedersen et al. i dette bind), har vi også foretaget opmærkning af 1500 sætninger fra samme korpus (DK-Clarín (dkclarin.ku.dk)) med et sæt af 58 grove betydningskategorier, de såkaldte supersenses der udgør en international standard, fx 'person', 'food', 'time', 'communication' etc., se Martínez et al. (2015) for flere detaljer. Et af de resultater der kom ud af opmærkningsarbejdet, er en statistik over hvilke af de 58 supersenses der er mest udbredte. Blandt de 6 hyppigste kategorier var substantiver og verber med supersense 'communication', og verber med supersense 'cognition' var syvendehyppigst. Ved at etablere et rammeleksikon der dækker de centrale rollebærende ord inden for netop disse to frekvente betydningsområder, regner vi med at der vil være forekomster nok i vores 1500 supersense-opmærkede sætninger til at rammeleksikonnet, og at vores hypotese i første omgang kan afprøves på de samme sætninger. Senere skal leksikonnet udvides til at omfatte i princippet alle betydningsområder i et nyt forskningsprojekt der er bevilget af Carlsbergfondet i 2016-2017, baseret på de resultater og metodiske erfaringer vi har opnået i indeværende forskningsprojekt.

Rammeleksikonnets leksikalske udgangspunkt, DDB, indeholder næsten 200.000 danske ord og udtryk (heraf 112.000 forskellige) sorteret i rækkefølge ud fra deres betydning og placeret under 888 afsnit og 22 kapitler med emneoverskrifter. I det grundmanuskript der ligger bag den trykte bog, er ordforrådet yderligere inddelt i 8.300 grupper opmærket med formaliserede semantiske oplysninger, og man kan derfor direkte udtrække de ca. 1700 grupper af verber og substantiver der beskriver en handling eller hændelse. Derved afgrænses netop det ordforråd som det er mest relevant at tildele BFN-rammer med henblik på at kunne udnytte det tilhørende engelske semantiske rolleinventar. Emneoverskrifterne i DDB gør det desuden muligt at identificere de ord der har kommunikations- eller kognitionsbetydning. Kapitel 11 "Tænkning" med 48 afsnit omfatter således kognition, mens kommunikationsordforrådet skal findes inden for de 49 afsnit i

kapitel 12 ”Sprog og kommunikation” samt i et udsnit af de i alt 58 afsnit i kapitel 15 ”Socialt liv” (fx i afsnit som ’Anerkendelse’, ’Smiger’ og ’Skælde ud’). En gruppe der er relevant for os, er fx gruppen af verber og substantiver der beskriver handlingen ’skælde ud’ i afsnit 15.24, se Figur 1.

```
{08_Vb_SbAflledning/has_hyperonym: 'vise sin vrede has_hyperonym: 'vredesudbrud involved_agent:
'person involved_patient: 'person}
> skælde ud, 'skrue bissen på, skænde, skælde, skælde (ud) for, 'tordne, 'tale dunder, tale med
store bogstaver, skælde og smælde, 'udskælde, 'gennemhegle, 'give (med) grovfilen, give en gang lak,
hegle igennem, 'sige et par borgerlige ord (til), give tørt på, skælde hæder og ære fra, skælde bælgen
fuld, skælde huden fuld, rive hovedet af nogen, 'tage nogen i skole, 'slå i bordet, 'bruge mund, 'herse,
overfuse, rise, dænge til<; > få luft for sin vrede, 'få afløb for sin vrede, 'bande nogen langt væk,
'rase ud<; > give luft for sin vrede, 'rase, 'fråde, 'se rødt, 'gå amok, 'springe/ryge i luften, 'koge over,
'eksplodere<; > snerre, 'bide ad, 'hvæse, 'spytte sætningen ud, 'sige vredt/bittert, 'råbe vredt<;
'komme efter nogen, 'småskænde på, 'vreden løb af med ham, gl'udøse sin vrede, 'skamme ud;
> 'komme med tilråb, 'fare i blækhuset, 'hvæsse/spidse pennen, 'hvæsse pennen<; 'hvæsse klørerne,
'forløbe sig; > vredesudbrud, raserianfald, udfald, 'vredesskrig, 'vris, 'snøft, 'gnaveri<;
> udskældning, 'skældud, 'udskæld, 'skænd, opsang, 'formaning, 'pegefinger, en sang fra de varme
lande uform<; > 'irettesættelse, 'røffel, 'gardinprædiken, 'moralprædiken eds<; > 'overhaling, 'skideballe
uform, 'sviner uform, '-et ordentligt pulver, 'møgfald, 'bredside, 'det glatte lag, 'balle, 'overfald, 'hak i tuden<;
> 'hårde ord, 'knubbede ord, 'salut, 'svada eds, 'salve, 'dundertale, 'tordentale, 'afskedssalut<;
'forløbelse; > skænderi, 'større skænderi, 'ophidset diskussion, 'hidsig diskussion, 'skændsmål gl<; >
{syn} 'heftigt skænderi, 'kæmpeskænderi<; 'familieskænderi
```

Figur 1: Gruppen af ord og udtryk med betydningen ’skælde ud’ i Den Danske Begrebsordbogs grundmanuskript.

De 1700 grupper i DDB indeholder i alt mere end 40.000 verber og verbalsubstantiver. Det høje tal skyldes et stort antal flerordsudtryk, herunder mange faste udtryk, men også en række udfoldninger af samme ord der i kombination med forskellige andre ord får netop den pågældende betydning, i Figur 1 fx *sige vredt/bittert* og *råbe vredt*. Desuden skyldes det gengangere på tværs af DDB’s manuskript, svarende til tildelingen af flere rammer til samme udtryk i BFN-modellen, jf. engelsk *talk* ovenfor. Men der er rent faktisk også en stor mængde forskellige ord, for ligesom ord med kommunikationsbetydning og kognitionsbetydning er hyppige i vores opmærkede tekster, er også DDB’s afsnit om disse betydningsområder meget ordrige. I alt udgør antallet af verber og verbalsubstantiver i vores udvalgte afsnit fra de tre kapitler, i alt 63 afsnit, ca. 7000 ord og udtryk, svarende til ca. 16 % af det samlede antal verber og verbalsubstantiver i begrebsordbogen.

3. Fremgangsmåde

Da alle ord og udtryk i DDB via fælles betydnings-id-numre i grundmanuskriptet er knyttet til deres betydningsbeskrivelse, valens-angivelser mm. i

DDO, kan de udtrukne lister af ord og udtryk fra DDB suppleres med valensmønstre til brug for præcis identifikation af hvilken handling eller situation der egentlig er tale om når man oversætter de danske ord til engelsk og skal finde ækvivalentens ramme i BFN. I overvejelserne indgår også både rammens andre leksikalske enheder og dens semantiske rolleinventar. I Figur 2 ses et udsnit af verber og verbaludtryk, igen fra begrebsområdet 'skælde ud', med tilknyttede valensmønstre fra DDO. Vi tager typisk udgangspunkt i det mest generelle udtryk i dansk når den rette ramme skal findes, fx for 'skælde ud'-gruppen verbet *skælde ud*, og vi anvender en digital tosprogsordbog for dansk-engelsk fra Gyldendal til oversættelsesarbejdet. Når en ækvivalent er fundet (*skælde ud = to scold*), søger vi efter det engelske ord *scold* i Lexical Unit-indekset i BFN og ser hvilken ramme det hører ind under. I tilfældet *scold* er det 'Judgment_direct_adress', og da beskrivelsen af den situation rammen dækker over, inkl. de roller den omfatter, passer fint til det vi forstår ved det danske begreb 'skælde ud', og også til de valensmønstre ord inden for begrebet typisk er beskrevet med i DDO, sættes denne ramme på både *skælde ud* og alle de nærsynonyme ord og udtryk fra DDB der passer på rammen.

Afsnit i DDB	Udtryk i DDB	Lemma i DDO	Valensmønster i DDO
15.24 Skælde ud	<i>skælde ud</i>	<i>skælde</i> , fast udtryk	NOGEN skælder ud på NOGEN NOGEN skælder NOGEN ud (for at../NOGET)
15.24 Skælde ud	<i>skrue bissen på</i>	<i>bisse</i> , fast udtryk	NOGEN skruer bissen på (over for NOGEN)
15.24 Skælde ud	<i>skænde</i>	<i>skænde</i> ., <i>bet. 1</i>	NOGEN skænder (på NOGEN) NOGEN skænder REPLIK
15.24 Skælde ud	<i>skælde</i>	<i>skælde</i> , fast udtryk	NOGEN skælder (på NOGEN)
15.24 Skælde ud	<i>tordne</i>	<i>tordne</i> , <i>bet. 2</i>	NOGEN tordner REPLIK NOGEN tordner mod NOGET/NOGEN
15.24 Skælde ud	<i>tale dunder</i>	<i>dunder</i> , fast udtryk	NOGEN taler dunder (til NOGEN)
15.24 Skælde ud	<i>skælde og smælde</i>	<i>skælde</i> , fast udtryk	NOGEN skælder og smælder (over NOGET/at../..)
15.24 Skælde ud	<i>gennemhegle</i>	<i>gennem-hegle</i> , <i>bet. 1</i>	NOGEN gennemhegler NOGEN

Figur 2: Ord og udtryk fra en handlingsverbalgruppe i afsnit 15.24 "Skælde ud" i DDB, suppleret med valensmønstre fra DDO.

4. Resultater og eksempler

I pilotstudiet fik de mere end 7000 verber og verbalsubstantiver på få måneders arbejde tildelt en rammebetegnelse fra BFN vha. den skitserede metode. I alt blev der anvendt 77 forskellige rammer af BFN's i alt 1073. Da de danske verber allerede er listet ud fra betydning og derfor typisk i én omgang kan oversættes til samme ramme, og da den rette ramme ofte findes inden for et stærkt reduceret antal af de over 1000 mulige i BFN fordi man bevæger sig inden for et overordnet betydningsområde (kommunikation eller kognition), var arbejdsmetoden meget effektiv. BFN har ikke defineret brede betydningsområder svarende til kapitlerne i begrebsordbogen, og man ville derfor ikke kunne indkredse hvilke rammer af de 1073 beskrevne der præcis dækker et bestemt betydningsområde som fx kommunikation. De rammer der er blevet anvendt ved oversættelse af de danske verber som vi som udgangspunkt definerede som hørende under emnet kommunikation i vores pilotstudie, blev dog typisk tildelt rammer fra BFN der umiddelbart også er genkendelige som hørende til dette tematiske område (fx 'Discussion', 'Become_silent', 'Bragging', 'Chatting' og 'Mention'). Inden for rolleinventaret i de anvendte rammer fandt man også mange fællestræk, fx omfatter de typisk en 'Speaker' eller en 'Communicator', men nogle af de anvendte rammer har et andet inventar, fx 'Interlocutor' (ved rammen 'Discussion') eller 'Authority' (ved rammen 'Deny_permission'). Det var dog også nødvendigt at benytte rammer der dækker mere end blot kommunikation, især til verber fra kapitel 15 'Socialt liv' i DDB, da afsnittene i dette kapitel ikke kun omfatter deciderede kommunikations-verber. Fx udgør *anerkende*, *værdsætte*, *skatte*, *agte* og *berømme* en ordgruppe i afsnittet 'Anerkendelse'. De fire første tildeles rammen 'Judgment'; kun det sidste, *berømme*, rammen 'Judgment_communication'. Som et eksempel på BFN's skelnen mellem rammer anføres det i rammebeskrivelsen for 'Judgment' at: "*This frame is distinct from the Judgment_communication frame in that this frame does not involve the Cognizer communicating his or her judgment to an Addressee*". Enkelte gange var der fuld overensstemmelse mellem et afsnits verber og en ramme i BFN. Fx svarer handlinger i afsnittet 'Fortale sig, sladre' i DDB ret præcist til rammen 'Reveal_secret' i BFN. I meget få tilfælde fandtes der ingen passende ramme i BFN; begrebet 'hemmeligholde' er fx (endnu) ikke dækket af nogen ramme. Vi har også noteret os enkelte rammer

i BFN der burde være anvendt i vores pilotstudie hvis man vil opnå fuld dækning af området kommunikation. Dette gælder fx 'Communication_means', og det skyldes at de danske verber står i et afsnit under et helt andet kapitel i DDB, nemlig det der beskriver apparater og teknik (i afsnit 19.17 'Telefon, e-mail'). Pilotstudiet giver således også feedback til arbejdet med DDB idet det er oplagt også at bringe ordene under kapitel 12 'Tegn, meddelelse, sprog'.

I langt de fleste tilfælde var der mange krydsninger mellem handlinger og hændelser i DDB's afsnit og rammer i BFN. Fx svarer det der kun udgøres af én ramme i BFN, 'Respond_to_proposal', til ord fra to afsnit i DDB, nemlig både 'Bekræfte' og 'Benægte'. Eller det omvendte var tilfældet: der er flere rammer i spil samtidig for en gruppe danske ord fra et afsnit i DDB, idet BFN afgrænser rammer fra hinanden ud fra betydningskriterier der ikke har været afgørende i DDB. Ved gruppen af danske 'skælde ud'-ord i DDB skelnes der fx ikke mellem at skælde ud direkte til en person eller skælde ud over vedkommende til en anden person. Men når BFN-rammerne skal tildeles de enkelte ord i afsnittet, skal man tage stilling til dette, dvs. om der er tale om rammen 'Judgment_direct_address' (der passer på de danske udtryk *skælde ud på nogen, skælde nogen ud, tale med store bogstaver, skælde nogen hæder og ære fra, give nogen tørt på, overfuse nogen, gennemhegle nogen, komme efter nogen, småskænde på nogen*), eller om der i stedet er tale om Judgment_communication hvor der ikke tales til den person man er vred på (fx *bande nogen langt væk*). Når det danske verbum har begge betydninger (evt. i forskellige konstruktioner: *rase mod nogen, rase over noget*), vælges den bredeste ramme 'Judgement_communication'. Til gengæld dækker både 'Judgment_communication' og 'Judgment_direct_adress' bredere mht. et andet betydningsaspekt end de modsvarende danske afsnit i DDB gør, idet der i BFN ikke skelnes mellem at bedømme positivt (fx *rose*) og bedømme negativt (fx *skælde ud*). BFN forudsætter i stedet at oplysninger ved selve ordene skal belyse om der er tale om negativ eller positiv bedømmelse. Figur 3 viser hvordan inddelingskriterier krydser hinanden i BFN og DDB.

DDB, afsnit BFN, ramme	15.19 Anerkendelse (positiv, både direkte og ikke direkte til person)	15.24 Skælde ud (negativ, både direkte og ikke direkte til person)
Judgement_communication (både positiv og negativ, ikke direkte til person)	<i>berømme</i>	<i>bande langt væk</i>
Judgement_direct_adress (både positiv og negativ, direkte til person)	<i>komplimentere</i>	<i>overfuse, gennemhegle</i>

Figur 3: BFN og DDB benytter forskellige kriterier i deres inddeling af ordforrådet.

4.1. Planer for brug af rammeleksikonet

Det rammeleksikon vi opbygger, vil som nævnt blive brugt til at forhåndsopmærke ord med rammer i vores 1500 supersenseopmærkede sætninger, vel at mærke i de tilfælde hvor ordet i korpusset er opmærket med supersense 'communication' eller 'cognition'. Ved den efterfølgende håndopmærkning af teksten vil annotørerne via oplysninger om den engelske rammes tilknyttede semantiske roller vide at de ved udtrykket *skælde ud* med rammen 'Judgment_direct_adress' skal lede efter og annotere en mulig 'Communicator', en 'Addressee', en 'Reason' eller et 'Topic' i den danske sætning. Vi har udvalgt et par eksempler på opmærkede sætninger (Figur 4).

Communicator	Addressee	Reason
Den ungarske landstræner	havde talt med store bogstaver til sine spillere	i pausen
Jeg skælder hende ud	for at være groft uansvarlig	
I debatten tordnes der løs	mod Det kgl. Teaters repertoire	

Figur 4: Semantiske roller for rammen 'Judgment_direct_adress' og eksempler på danske, opmærkede sætninger (i form af citater fra DDO).

Leksikonet indeholder langt flere ord og udtryk til hver ramme end man typisk finder i BFN, særligt flere flerordsudtryk. Nogle er kontinuerte, andre ikke, nogle kan varieres, andre ikke. I det kommende annoteringsarbejde vil det vise sig om flerordsudtrykkene rent faktisk kan genkendes i den automa-

tiske morfologiske analyse og dermed også udnyttes i den automatiske præannotering. Flerordsudtrykkene vil også kræve særlig opmærksomhed fra annotørernes side idet det semantiske rolleinventar der beskrives for rammen i BFN, mange gange vil være at finde i udtrykket selv i det danske leksikon fordi det af formidlingshensyn i DDB er nævnt i form af et pronomen eller en typisk leksikalsk udfyldning (fx i *skælde ud på nogen* og *skælde nogen ud*). I andre tilfælde er det i stedet tale om et fast udtryk, dette gælder fx *tale med store bogstaver*; her skal 'med store bogstaver' ikke tildeles nogen rolle i opmærkningen.

5. Konklusion

Pilotstudiet har bekræftet vores hypotese om at man med fordel kan kombinere DDB's inddeling af verber og verbalsubstantiver i emnegrupper og FrameNets rammer. De to semantiske ressourcer har begge en tematisk inddeling af handlinger og hændelser som noget meget centralt i beskrivelsen af ordforrådet, og det betyder dels at man kan koble store grupper af verber til en passende ramme i én arbejdsgang, dels at man kan fokusere på ordforrådet inden for et bestemt betydningsområde hvilket letter arbejdet betydeligt. Men man kan ikke overraskende konstatere at de lægger snittet mellem emnerne lidt forskelligt. Nogle gange omfatter en ramme fra BFN derfor verber og verbalsubstantiver fra flere afsnit i DDB, nogle gange kun en delmængde af verberne og verbalsubstantiverne i ét afsnit. Der er for så vidt ikke en af inddelingene der er mere korrekt en den anden, men de rummer supplerende typer af information der med fordel kan bringes sammen. Den konkrete anvendelse af leksikonnet i vores kommende opmærkningsarbejde vil vise om vores hypotese om at de engelske rollebeskrivelser også passer for dansk, holder, eller om vi i stedet er nødsaget til at udvide det danske rammeleksikon vi opbygger på basis af DDB's ordforråd, med roller specifikt for dansk.

Litteratur

Ordbøger

BFN = *Berkeley FrameNet*: <framenet.icsi.berkeley.edu>.

DDB (2014) = S. Nimb, H. Lorentzen, L. Theilgaard & T. Troelsgård: *Den Danske Begrebsordbog*. København: Det Danske Sprog- og Litteraturselskab.

DDO = ordnet.dk/ddo – E. Hjort, K. Kristensen et al. (2003-2005): *Den Danske Ordbog*. København: Det Danske Sprog- og Litteraturselskab.

Anden litteratur

Bick, Eckhard (2011): A FrameNet for Danish. I: B. S. Pedersen, G. Nešpore & I. Skadina (red.): *Proceedings of NODALIDA 2011*. Tartu, 34-41.

Boas, Hans C. (red.) (2009): *Multilingual FrameNets in computational lexicography: Methods and applications*. Trends in linguistics, Studies and Monographs 200. New York/Berlin: Mouton de Gruyter.

Borin, Lars & Markus Forsberg (2009): All in the Family: A Comparison of SALDO and WordNet. I: *Proceedings of the NODALIDA 2009 Workshop on WordNets and other Lexical Semantic Resources - between Lexical Semantics, Lexicography, Terminology and Formal Ontologies*. NEALT Proceedings Series. Göteborg: Språkbanken, Göteborgs Universitet, 7-12.

Friberg, Karin Heppin & Dana Dannells (2015): Polysemy and questions of lumping or splitting in the construction of Swedish FrameNet. I: *Proceedings of the Workshop on Semantic resources and Semantic Annotation for Natural Language Processing and the Digital Humanities*. Linköping: Linköping University Electronic Press, 12-20.

Martinez, Hector, Anders Johannsen, Sussi Olsen, Sanni Nimb, Nicolai Hartvig Sørensen, Anna Braasch, Anders Sjøgaard & Bolette Sandford Pedersen (2015): Supersense tagging for Danish. I: *Proceedings of the 20th Nordic Conference on Computational Linguistics*. NODALIDA 2015. Linköping: Linköping University Electronic Press, 21-29.

Sanni Nimb
seniorredaktør
Det Danske Sprog- og Litteraturselskab
Chr. Brygge 1
DK-1219 København K
sn@dsl.dk

Bolette S. Pedersen
professor, institutleder
Nordisk Forskningsinstitut
Københavns Universitet
Njalsgade 136, bygn. 27. 2
DK-2300 København S
bspedersen@hum.ku.dk

Betydningsinventarer – i ordbøger og i løbende tekst

Bolette Sandford Pedersen, Sanni Nimb, Anna Braasch & Sussi Olsen

We examine a set of highly polysemous nouns in Danish in order to understand how well word senses in Den Danske Ordbog (DDO) and the Danish wordnet, DanNet, correlate with the meanings found in examples of running text. The overall goal of the task is to provide adequate training data for automatic word sense disambiguation. To this end, we annotate a corpus with a full sense inventory from DDO and compare with the annotations provided on the basis of a *clustered* inventory derived automatically from DDO and DanNet's ontological classes. The results show that our hypothesis of clustered senses being more manageable and providing higher annotation agreement (apart from what can be expected from a higher chance agreement when there are less annotation options), holds for a vast majority of the studied cases. Further, the investigations provide a valuable assessment of the organization of senses in the studied dictionaries.

1. Et betydningsopmærket korpus – hvorfor?

I forbindelse med et samarbejdsprojekt mellem Københavns Universitet og Det Danske Sprog- og Litteraturselskab¹ betydningsopmærker vi en offentligt tilgængelig korpusressource med betydningsdistinktioner i dansk. Det praktiske, sprogteknologiske formål med denne opmærkning er at skabe sprogresourcer der muliggør træning af computermodeller til at kunne entydiggøre ord på dansk og derigennem forbedre teknologiske sprog-tjenester som søgemaskiner, maskinoversættelse og spørgsmål-svarsystemer.

Imidlertid har vi under opmærkningen særligt fokus på at udforske forholdet mellem de betydningssskel der etableres i den konventionelle ordbog, og de betydningssskel som man intuitivt kan genkende og opnå konsensus om når man opmærker betydninger i løbende tekst. For at belyse dette område

¹ Projektet "Semantic Processing across Domains" finansieres af Forskningsrådet for Kultur og Kommunikation i perioden 2013-2016 (jf. Pedersen et al. 2014 og <cst.ku.dk/projekter/semantikprojekt/>).

bedst muligt er alle de korpusopmærkninger vi diskuterer i artiklen, opmærket af flere annotører, og enigheden er beregnet.

Vi opmærker korpusset med betydningsinventarer af varierende grovhed. Disse betydningsinventarer fremkommer ved systematisk at kombinere den hierarkiske struktur fra en konventionel ordbog (Den Danske Ordbog; DDO, jf. ordnet.dk) bestående af hoved- og underbetydninger med den ontologiske viden der er indarbejdet i betydningerne i det leksikalsk-semanticke wordnet, DanNet (Pedersen et al. 2009, jf. www.andreord.dk) i form af typer som HUMAN, EVENT og PROPERTY. Ved at sammenligne annoteringerne med et grovkornet betydningsinventar med annoteringer med et mere finkornet, prøver vi at komme nærmere en forståelse af hvor det er svært at skelne for mennesker – og hvor det derfor sandsynligvis også er svært for en computermodel. Helt overordnet er målet at nærme os et ”passende” niveau af finkornethed; et som er håndterbart for de menneskelige annotører og for de maskinlæringsystemer der efterfølgende skal processere dem, men som samtidig er tilstrækkelig nuanceret til at kunne forbedre eksisterende teknologiske sprog tjenester.

Undervejs i denne proces opnås i tillæg en større erkendelse af hvilke grundlæggende problemstillinger man står overfor i det leksikografiske arbejde når man skal lægge sig fast på hvilke principper man skal anvende i beskrivelsen af betydningsinventaret. Vi indleder i afsnit 2 med en beskrivelse af hhv. DDO's og DanNets betydningsstruktur og redegør på baggrund heraf for hvordan vi automatisk kan udforme klynger af betydninger. Endelig redegør vi for den udvalgte empiri i form af 20 meget polyseme substantiver. I afsnit 3 går vi i dybden med annoteringsarbejdet og beskriver nogle af de divergenser der opstår mellem annotørerne. I afsnit 4 evaluerer vi det samlede annoteringsarbejde.

2. Betydningsstruktur og etablering af betydningsklynger

Inden vi omtaler undersøgelsen, vil vi kort fremlægge vores hypoteser ud fra teoretiske beskrivelser af systematiske sammenhænge mellem bestemte typer af betydning af samme ord, og den beskrivelse af ordforrådet man finder i DDO og DanNet. Cruse (2000:111) introducerer flere typer af relationer der beskriver sammenhæng mellem betydninger af samme ord (modsat semantiske relationer der beskriver sammenhæng mellem forskellige ord, fx de rela-

tionstyper der anvendes i wordnets). Disse ”interne” relationer er oplagte at tage udgangspunkt i når man ønsker at sammenlægge betydninger i en ordbog med henblik på at opnå et mindre detaljeret betydningssinventar. Autohyponymi kalder han fx den systematiske sammenhæng der er indenfor ord der både har en bred og en mere indsnævret betydning med samme overbegreb (fx *drikke* (væske) over for *drikke* (for meget alkohol)). Det modsatte fænomen, sammenhængen fra en snæver til en udvidet betydning af samme ord, og hvor de to betydninger har samme overbegreb, betegner han autosuperordination (fx *mand* (person af hankøn) over for *mand* (person uden tanke på køn)). Hvis sammenhængen mellem et ords to betydninger bygger på en helhed set i forhold til en del af samme helhed, bruger han to termer: automeronymi (del for helhed, fx *dør* i betydningen ’åbning’ hhv. ’plade’) og autoholonymi (helhed for del, fx *krop* i betydningen ’hele legemet’ over for *krop* i betydningen ’kun den centrale del af kroppen’, ’torsoen’).

I DDO's struktur afspejles de systematiske sammenhænge inden for samme ord som Cruse observerer og navngiver, i princippet via ordbogens hierarkiske organisering i hoved- og underbetydninger. *Krop* = ’legeme’ er fx hovedbetydning (overbetydning) til underbetydningen *krop* = ’torso’ i DDO. Dette fremgår også af DDO's brugervejledning der fastslår at en underbetydning kan være enten ”en faglig betydning, en delbetydning, en overført betydning eller en udvidet betydning i forhold til overbetydningen” (DDO, bind 1, s. 27). Da en faglig betydning vil være en indsnævret betydning med samme overbegreb som overbetydningen, er relationen mellem dem af typen autohyponymi. Relationen mellem en delbetydning og en overbetydning svarer til automeronymi. Det som i DDO betegner udvidet betydning, kan manifestere sig både i form af autosuperordination (samme overbegreb som overbetydningen, men bredere betydning) og i form af autoholonymi (overbetydningen betegner en del af noget, men underbetydningen er bredere og betegner helheden).

Man finder dog også eksempler på at de forskellige sammenhænge er usynliggjort i DDO's struktur. Substantivet *dør* er fx beskrevet med to hovedbetydninger selv om der er en automeronymirelation mellem dem, det samme er de to betydninger af verbet *drikke*, selv om der er en autohyponymirelation mellem dem. Og substantivet *mand* (person af hankøn), *mand* (i den

indsnævrede betydning 'ægtemand = autohyponymi') og *mand* i den udvidede betydning 'person/menneske uden tanke på køn' (autosuperordination) er beskrevet som tre hovedbetydninger og ikke som en overbetydning med to underbetydninger. Dette skyldes at DDO-redaktionen af formidlingsgrunde foretrak en flad og let tilgængelig struktur når det følte logisk, fx når den betydning der principielt skulle have været en underbetydning, var mest frekvent i sproget. En yderligere, ikke-ubetydelig faktor er at der i leksikografisk arbejde altid er en generel usikkerhedsfaktor mht. hvor langt et udvidet eller indsnævret betydningsaspekt må fjerne sig fra sin kernebetydning før det bør udløse en ny hovedbetydning (jf. Svensén 2009:212).

DanNet er et wordnet for dansk der er baseret på DDO's betydningsinventar og de semantiske relationer mellem de forskellige ords betydninger, og som samtidig grupperer synonyme og nærsynonyme betydninger i såkaldte synsets. I DanNet kan man derimod intet udlede om relationen mellem et enkeltords flere betydninger, ej heller hvilken af de måske mange betydninger der er mest prominent, idet alle betydninger ligestilles. Til gengæld er hvert synset i DanNet forsynet med en ontologisk type baseret på EuroWordNets topontologi (Vossen et al. 1999), som er en hierarkisk ontologi der består af ca. 40 kategorier af typen HUMAN, EVENT, PROPERTY, DISEASE, ARTIFACT, BUILDING mv. Denne oplysningstype danner sammen med DDO's hoved- og underbetydninger udgangspunktet for en automatisk klyngedannelse på et velunderbygget grundlag.

På trods af at DDO's struktur langt fra i alle tilfælde afspejler de systematiske sammenhænge der er mellem betydninger, har vi alligevel en vis forventning om at især en stor del af de indsnævrede betydninger er beskrevet som underbetydninger der via viden om et fælles overbegreb (udtrykt via DanNets ontologi) med god mening kan slås sammen med deres hovedbetydning til én betydning. Man finder fx substantivet *område* i DDO beskrevet med en hovedbetydning 'landområde' (LOCATION) og en indsnævret underbetydning 'administrativt landområde'; disse to betydninger kan lægges sammen automatisk via DanNets ontologiske oplysninger, hvilket intuitivt giver god mening. Vi forventer også en del automatiske sammenlægninger i de tilfælde hvor del/helhedsrelationer er afspejlet i den hierarkiske struktur i DDO og samtidig kan identificeres via DanNets ontologiske oplysninger om betydningerne. Fx beskrives substantivet *sten* i DDO med hovedbetydningen 'fast, hårdt materiale' og derefter med to underbetydninger 'stykke af et

sådant materiale' og 'stykke der anvendes som byggemateriale'; her vil alle tre betydninger blive slået sammen til kun én via den autoholonymi-relation der kan udledes af DanNets ontologiske typer, hvilket umiddelbart også giver god mening.

Ud fra et ønske om at arbejde med et så stort antal sammenlagte betydninger som overhovedet muligt udvalgte vi til vores undersøgelse et antal substantiver med mange betydninger der samtidig i høj grad var hierarkisk organiseret, dvs. at antallet af underbetydninger skulle være relativt højt i forhold til antallet af hovedbetydninger. 20 substantiver blev udvalgt på dette grundlag idet det samtidig var et krav at de havde en høj forekomst i tekster, målt i frekvens i det korpus som dannede grundlaget for beskrivelsen af lemmerne i DDO, nemlig Korpus90. På den måde sikrede vi os at korpusset vi anvender, sandsynligvis ville indeholde det nødvendige antal teksteksempler der skulle bruges i opmærkningsarbejdet beregnet ud fra en generel tommelfingerregel om at man bør opmærke 100 eksempler + 15 x antal betydninger for et givent lemma (jf. fremgangsmåden anvendt i Senseval-projektet: www.senseval.org).

Efter sammenlægningen af betydninger ud fra de ontologiske typer der var tildelt i DanNet, opnåede vi at 18 af de 20 substantiver fik reduceret deres antal af betydninger, se Tabel 1; gennemsnitligt set reduceres antallet af betydninger med 23,5 procent.

	Antal betydninger i DDO	Antal klynger
selskab	10	6
plads	13	9
slag	17	11
plade	11	10
skud	12	12
ansigt	7	6
skade	6	5
stykke	18	16
kontakt	9	5
stand	9	7
top	8	6
kort	10	4
vold	9	7
hul	14	11
hold	12	10
lys	13	9

blik	7	6
model	8	7
kurs	3	3
tang	Udeladt pga. for få forekomster i korpus	

Tabel 1: Udvalgt i empiri: 20 polyseme substantiver og deres betydningsreduktioner – i snit 23,5 % reduktion.

Som illustration af hvordan klyngedannelsen kommer til udtryk i de enkelte leksikalske indgange, ses den leksikalske indgang for *selskab* i Figur 1 hvor betydningerne 1, 1a og 1b klynges sammen idet de har den samme ontologiske type i DanNet: RELATION. Underbetydning 1c udgør derimod sin egen klynge da den har den ontologiske type PERSON. Hovedbetydning 2 er også af typen PERSON, men udgør igen sin egen klynge ud fra princippet om at hovedbetydninger aldrig klynges sammen. Hovedbetydning 3 er af typen EVENT (altså en begivenhed). Hovedbetydning 4 og underbetydning 4a giver igen anledning til en klynge da typen her er INSTITUTION, hvorimod hovedbetydning 5 udgør sin egen klynge. De i alt 10 betydninger reduceres altså hermed til 6 klynger. Herudover følger en række faste udtryk som alle bibeholdes og udgør hver deres egen klynge.

HOVEDBETYDNING 1

det at være eller foretage sig noget sammen med en eller flere andre personer

1a. samvær eller fællesskab med en eller flere andre, ofte af underholdende eller adspredende karakter

1b. det at noget forekommer sammen med noget andet

1c. personkreds eller miljø som man knytter sig til eller bliver forbundet med af personlige eller professionelle årsager

HOVEDBETYDNING 2

gruppe af personer som foretager sig noget bestemt i fællesskab

2a. gruppe af personer som er samlet til en fælles social aktivitet, fx en fest el. middag

HOVEDBETYDNING 3

festlig sammenkomst med mange mennesker, mad og drikke og evt. underholdning

HOVEDBETYDNING 4

handels-, industri- eller erhvervsvirksomhed

4a. sammenslutning mellem to eller flere juridiske personer som har til formål at drive en handels-, industri- eller anden virksomhed fx et aktieselskab eller et anpartsselskab

HOVEDBETYDNING 5

sammenslutning af personer med en fælles, oftest faglig eller akademisk interesse

Figur 1: *Selskabs* betydninger i DDO.

3. Om annoteringsarbejdet

De tekster vi opmærker, er udtrukket fra Det Danske CLARIN Reference Corpus (Asmussen 2012) som er et almensprogligt korpus på 45 mio. løbende ord, med hovedvægten på avistekster (48 %), men korpuset omfatter også mange andre forskellige teksttyper. Vi har lagt vægt på at have et bredt udsnit af teksttyper repræsenteret i vores korpus da en optimering af sprogprocessering *på tværs* af teksttyper er et vigtigt mål for vores projekt. I Figur 2 kan man se en liste over de teksttyper vi har opmærket.

Titel	Beskrivelse
<i>Bentes blog</i>	En blog skrevet af en kvinde i fyrrerne
<i>Selvhenter</i>	Et chatforum for især unge mennesker
<i>Mangamania</i>	Et chatforum for mangafans
<i>Se og Hør</i>	Ugeblad
<i>Folketingstaler</i>	Nedskrevne folketingsdebatter
<i>Politiken</i>	Avis

Figur 2: Teksttyper der opmærkes.

For hvert ord udvælges et antal sætninger som beskrevet i afsnit 2.4 Eksempelvis annoteres for *selskab* $100 + 15 \times 7 = 205$ teksteksempler. Vi har ved udvælgelsen af sætninger valgt at lægge særlig vægt på at få repræsenteret de sociale medier, og alle sætninger indeholdende de aktuelle ord medtages fra disse teksttyper. Det faktum at der er vis uoverensstemmelse mellem vores valgte korpusmateriale og det korpus som DDO-betydningsbeskrivelserne er baseret på, giver nogle problemer. DDO's interne korpus på 40 mio. tokens (svarende nogenlunde til Korpus90, jf. Asmussen (2006)) er for det første ældre, for det andet var det vægtet bredere hvad angår teksttyper end Clarin-korpuset (først og fremmest i form af langt flere litterære tekster), for det tredje indeholdt det af gode grunde ikke tekster fra sociale medier. Vi mener dog at fællesnævneren i form af relativt meget avisstof i begge korpura til en vis grad udjævner de forskelle der er. Fx optræder de polyseme substantiver, vi opmærker i denne sammenhæng, og som er velrepræsenterede i Korpus90, ikke særligt hyppigt i tekster fra de sociale medier, så til trods for denne vægtning er især avistekster og folketingsdebatter statistisk set i overtal i vores korpuser.

Annoteringen af de udvalgte substantiver foretages både som *finkornet opmærkning*, dvs. med det fulde, detaljerede DDO-betydningsinventar og som *klyngeopmærkning*, dvs. med bredere betydningsklynger der er dannet ved sammenlægning af betydninger (jf. ovenfor, afsnit 2). Formålet er at undersøge hvorvidt den finkornede eller den klyngebaserede opmærkningsmetode giver den højeste annotørenighed, idet graden af annotørenighed er en rimelig god målestok for metodens anvendelighed.

En gruppe på tre studerende har stået for opmærkningen af teksterne. Derudover har flere af forskerne i projektet opmærket et antal filer. Alle tekster er blevet opmærket af mindst to personer, så det er muligt at måle enigheden mellem annotørerne og undersøge de tilfælde hvor de divergerer. Teksterne opmærkes i opmærkningsværktøjet WebAnno (se Figur 3), et webbaseret værktøj udviklet af Technische Universität Darmstadt for CLARIN (Yimam et al. 2013). WebAnno kan tilpasses mange opmærkningstyper og gør det let at bevare overblikket over fremdriften, lave en korrigeret version af de anoterede filer samt ikke mindst beregne kvaliteten af opmærkningerne målt ud fra enigheden annotørerne imellem.



Figur 3: Annotering i WebAnno, *selskab* med klyngeinventar.

Der synes generelt at være to grundlæggende forhold der volder problemer for annotørerne, og som kommer til udtryk som uenigheder i annoteringerne:

- graden af *kompleksitet* i et givet ords betydningsstruktur (også afspejlet i dets reducerede betydningsklynger), og
- antallet og arten af de *faste udtryk*.

Hvis vi indledningsvis ser på kompleksitet i betydningsstruktur, kan vi se at annoteringen generelt forløber uproblematisk når der er tale om en simpel betydningsstruktur med dels klare skel mellem (få) hovedbetydninger og

dels relativt få underbetydninger. Et godt eksempel herpå er *model*, idet ordet i DDO har seks klart adskillelige hovedbetydninger (med i alt to underbetydninger) og et enkelt fast udtryk (*stå model til*); disse udgør det finkornede opmærkningsinventar. De i alt otte finkornede hoved- og underbetydninger er slået sammen til syv betydningsklynger, hvilket selvsagt er en ganske lille reduktion; det faste udtryk følger uændret med. Hver af de to opmærkningsmetoder resulterer i relativ høj grad af annotørenighed (hhv. 0,66 og 0,76 i henhold til Krippendorffs α^2 ; se også afsnit 4). Hypotesen om at et mindre detaljeret inventar giver en mere konsistent opmærkning, bekræftes altså her.

Opmærkningen af *kontakt* viser et mere sammensat mønster: Ved konkrete betydninger med klare betydningskæder er der stor enighed, som i eksempel 1:

- 1) *Afblegning af hår må kun finde sted i striber eller så afblegningsmidlet ikke er i **kontakt** med huden*

Sætninger af denne type er af alle opmærket med betydningsklyngen **kontakt_2_2a** → ”det at to genstande, dele, flader el.lign. er i fysisk forbindelse eller berøring med hinanden”. Ligeledes i eksempel 2:

- 2) *For selv om det er bedst at slukke på **kontakten** for ting, du ikke bruger, så efterlever de færreste rådet fuldt ud*

Sætninger af denne type er af alle opmærket med **kontakt_3_3a** → ”teknisk indretning som [...]giver mulighed for at tilslutte og afbryde en elektrisk strøm i en ledning og dermed tænde eller slukke for noget”.

Men enigheden ophører så snart vi bevæger os over i abstrakte eller overførte betydninger, hvor skellet er utydeligt, især hvis ordets kontekst er kort og/eller kan fortolkes på forskellig vis, som vist i eksempel 3:

- 3) *Gennem hårdt arbejde og omfattende kontakter fik Udenrigsministeriet lokaliseret danskeren og fik konsulær adgang*

² Vi anvender Krippendorffs α til beregning af annotørenighed som bl.a. modregner det faktum at det alt andet lige er nemmere at opnå enighed om få betydninger end om mange (Krippendorff 2011).

Her opstår der uenighed mellem klyngerne **kontakt_1_1a** → ”forhold mellem to parter som indebærer indbyrdes kommunikation”, **kontakt_1b** → ”person eller part som man har forbindelse med eller kommer i berøring med fagligt”, og **kontakt_1c_1_d** → ”det at sætte sig i forbindelse med en anden part, ofte for at indlede et samarbejde eller for at opnå noget eftertragtet; OVERFØRT det at der er forbindelse og dermed grundlag for kommunikation, indsigt”.

Meget ofte er der desuden delvis enighed om at et ord er brugt i overført betydning, men hvilken af dem den enkelte annotør vælger, synes lidt tilfældigt, som belyst i eksempel 4.

- 4) *I dette lys er det ikke betryggende at opleve repræsentanter [...] tage afstand fra menneskerettighederne og ligefrem tale om[....]*

Eksemplet er opmærket med hhv. **lys_1f** → ”OVERFØRT opmærksomhed, herunder: søgelys, rampelys” (DDO), med **lys_1e** → ”OVERFØRT viden; (religiøs) indsigt, herunder: forklarelsens lys, erkendelsens lys” (DDO), og endelig med **lys_F_i-et-nyt-(andet, -...)-lys** → som (variant af) fast udtryk (listet i DDO).

Opmærkning af *faste vendinger* burde i princippet være helt uproblematisk da man blot skal forholde sig slavisk til den liste af vendinger som er opstillet i DDO. Dette forhold ses også meget tydeligt i opmærkningen af *stand*: dels er enigheden meget høj med både fuldt og klyngebaseret inventar, dels nås næsten den samme α -værdi ved de to metoder. Dette skyldes at langt størstedelen af eksemplerne (79 af 100) udgøres af faste udtryk af typen *i stand til* og *ude af stand til*, og kun 16 forekomster er opmærket med egentlig betydning (fordelt på fire betydninger). Der er dog nogen usikkerhed/uenighed hos annotørerne med hensyn til faste udtryk når 1) DDO's liste mangler et etableret og hyppigt forekommende fast udtryk, 2) det faste udtryk i korpus er en syntaktisk eller leksikalsk variant af et udtryk på DDO's liste (med semantisk nærliggende betydning), 3) korpusforekomsten er en kontamination af to faste udtryk, og endelig 4) der er syntaktisk og leksikalsk sammenfald mellem ordets forekomst i et kompositionelt udtryk og i et fast udtryk.

4. Evaluering og konklusion

Den overordnede konklusion på annoteringeksperimentet kan sammenfattes til at en ontologisk baseret klyngetilgang til betydningsinventaret, hvor man opmærker med klynger af betydninger, opnår en bedre enighed mellem annotørerne beregnet ved Krippendorffs α end opmærkning med det fulde DDO-inventar. De enkelte polyseme ords resultater kan aflæses i Tabel 2. Således fremgår det at hypotesen bekræftes i 67 % af tilfældene og afkræftes i 22 % af tilfældene, mens 11 % af de polyseme ord fremviser samme enighed med uden og klyngeannotering.

	Enighed v. finkornet opmærkning	Enighed v. klyngeopmærkning	Bekræftelse af hypotese
selskab	0,48	0,81	Ja
plads	0,51	0,63	Ja
slag	0,50	0,51	Ja
plade	0,048	0,051	Ja
skud	0,29	Ingen reduktion	Neutral
ansigt	0,33	0,35	Ja
skade	0,59	0,65	Ja
stykke	0,55	0,62	Ja
kontakt	0,50	0,60	Ja
stand	0,80	0,86	Ja
top	0,57	0,46	Nej
kort	0,48	0,51	Ja
vold	0,38	0,57	Ja
hul	0,52	0,42	Nej
hold	0,64	0,60	Nej
lys	0,64	0,56	Nej
blik	0,45	0,53	Ja
model	0,66	0,76	Ja
kurs	0,84	Ingen reduktion	Neutral
tang	For få belæg		

Tabel 2: Samlet evaluering af de udvalgte polyseme substantiver beregnet med Krippendorffs α .

Men som det også fremgår af de udvalgte eksempler i afsnit 3, forekommer der meget store udsving fra det ene polyseme ord til det andet, og antallet af betydninger synes ikke i sig selv at være afgørende for hvor vanskeligt det er at opnå annotørenighed. Ikke overraskende udgør ord med mange abstrakte og metaforiske betydninger et problem både med det finkornede betyd-

ningsinventar og med det klyngebaserede. I flere tilfælde kunne det faktisk se ud til at det udvalgte ordforråd har fået nye betydningsnuancer siden det blev registreret i DDO ud fra et snart 25 år gammelt korpus. Det gælder især den udbredte brug af metaforisk sprog, samt forholdet til hvilke af de faste udtryk der er gængse og hyppige.

I vores videre arbejde kunne vi tænke os at kaste et mere kritisk blik på hovedbetydningernes status i forhold til underbetydningerne, jf. diskussionen i afsnit 2. Det kunne fx være interessant at belyse om en sammenlægning af hovedbetydninger med samme ontologiske type generelt vil føre til for stort informationstab, eller om det faktisk vil styrke ressourcen i forhold til opmærkning og senere automatisk entydiggørelse. Et sådant undersøgelsesarbejde ville i tillæg kaste yderligere lys over konsistensniveauet i de anvendte leksikografiske data.

Tak

Tak til studenterannotørerne Sarah Lee Naldal, Selma Rosenfeldt-Olsen, Ida Hauerberg Wolthers, samt til Nicolai Hartvig Sørensen og Héctor Martínez Alonso for teknisk støtte til korpus- og annotationsarbejdet.

Litteratur

- Asmussen, J. (2006): Towards a methodology for corpus-based studies of linguistic change. Contrastive observations and their possible diachronic interpretations in the Korpus 2000 and Korpus 90 Corpora of Danish. I: Dawn Archer/Paul Rayson/Wilson (eds.): *Corpus Linguistics Around the World*. Amsterdam: Rodopi, 33-48.
- Asmussen, J. (2012): *CLARIN-Referencekorpus*. Foredrag ved Sprogteknologisk Workshop, Københavns Universitet, 31. oktober 2012. <cst.ku.dk/Workshop311012/sprogtekno2012.pdf>.
- Cruse, D. A. (2000): *Meaning in language*. Oxford: Oxford University Press.
- DDO = *Den Danske Ordbog* (E. Hjorth et al.) 2003-2005. Det Danske Sprog- og Litteraturselskab & Gyldendal: København.

- Fersøe, H. B. Maegaard & B. S. Pedersen (2013): Humanities eInfrastructure initiatives in Denmark. I: *Proceedings of the workshop on Nordic language research infrastructure at NODALIDA 2013*. Linköping University Electronic Press, Vol. 089.
- Pedersen, B.S., S. Nimb, S. Olsen, A. Søgaard & N. Sørensen (2014): Semantic Annotation of the Danish CLARIN Reference Corpus. I: *Proceedings from isa-10, 10th Joint ACL - ISO Workshop on Interoperable Semantic Annotation*, Reykjavík, 25-29.
- Pedersen, B.S., S. Nimb, J. Asmussen, N. Sørensen, L. Trap-Jensen & H. Lorentzen (2009): DanNet – the challenge of compiling a WordNet for Danish by reusing a monolingual dictionary. I: *Language Resources and Evaluation, Computational Linguistics Series* 43(3), 269-299.
- Krippendorff, K. (2011): Agreement and information in the reliability of coding. I: *Communication Methods and Measures* 5(2), 93-112.
- Svensén, B. (2009): *A Handbook of Lexicography*. Cambridge: CUP.
- Vossen, P. (ed.) (1999): *EuroWordNet, A Multilingual Database with Lexical Semantic Networks*. Netherlands: Kluwer Academic Publishers.
- Yimam, S.M., I Gurevych, R. Eckart de Castilho & C. Biemann (2013): WebAnno: A Flexible, Web-based and Visually Supported System for Distributed Annotations. I: *Proceedings of ACL-2013*, demo session, Sofia, Bulgaria.

Bolette S. Pedersen
professor, institutleder
Nordisk Forskningsinstitut
Københavns Universitet
Njalsgade 136, bygn. 27. 2
DK-2300 København S
bspedersen@hum.ku.dk

Sanni Nimb
seniorredaktør
Det Danske Sprog- og Litteraturselskab
Chr. Brygge 1
DK-1219 København K
sn@dsl.dk

Anna Braasch
seniorforsker emeritus
Center for Sprogteknologi, NFI
Københavns Universitet
Njalsgade 140, bygn. 25.4.
DK-2300 København S
braasch@hum.ku.dk

Sussi A. Olsen
videnskabelig medarbejder
Center for Sprogteknologi, NFI
Københavns Universitet
Njalsgade 140, bygn. 25.4
DK-2300 København S
saolsen@hum.ku.dk

Rietz och riksspråket

Eva Thelin

The most recent dialect dictionary covering all Swedish dialects appeared in 1862–67, published by Johan Ernst Rietz. In those days, dialect dictionaries only consisted of words which did not appear in standard Swedish. Rietz, though, does include some standard words and in the foreword to the dictionary he specifies his criteria for entering them. In this article I examine what standard words Rietz has actually included (based on the words beginning with h-) and how well they correspond to his criteria.

1. Bakgrund

1862–67 utkom Svenskt dialekt-lexikon eller Ordbok öfver svenska allmogespråket (i fortsättningen kallat DL). DL omfattar knappt 860 sidor och söker beskriva alla svenska dialekter, även finlandssvenska och estlandssvenska, och det är än idag den senast utgivna ordboken med den inriktningen. DL framställdes av Johan Ernst Rietz (1815–1868), vid tiden kontraktsprost i Oxie i Skåne.¹ Han hade publicerat en hel del dessförinnan, framför allt utgåvor av fornsvenska texter och under åtminstone 17 år var han sysselsatt med en fornsvensk ordbok. Under arbetet med den lade han märke till att många av de fornsvenska orden fanns bevarade i dialekterna – en observation han inte var den förste att göra. Ett intresse för dialekterna väcktes hos Rietz, vilket så småningom ledde till att han lade det fornsvenska ordboksmanuskriptet åt sidan och övergick till att utarbeta en dialektordbok. Hur Rietz gick till väga vid framställandet av DL redogör Vidar Reinhammar närmare för i artikeln Om J. E. Rietz och Svenskt dialekt-lexikon (1998), vilket är den enda undersökning som hittills utförts om lexikonet. Reinhammar, som på sin tid ledde projektet Ordbok över Sveriges dialekter (OSD), hade haft anledning att granska och jämföra DL

¹ Detaljer om Rietz biografi kan inhämtas ur bl.a. Carlquist (1948), Ekholm (2001) och Svenskt biografiskt lexikon.

med dess källmaterial inför att de skulle tillföras OSD:s samlingar. Han uppdagade då ett antal felaktigheter i DL, vilka redovisas i artikeln.

Sedan 2012 arbetar jag på en doktorsavhandling om DL, vilken inriktar sig på att undersöka lexikonets strukturer på mikro- och makronivå samt att studera vilka principer Rietz tillämpat vid lemmaselektionen.

2. Lemmaselektion i dialektordböcker

I denna artikel inriktar jag mig på Rietz principer för lemmaselektion, närmare bestämt hur han förhållit sig till de ord som dialekter och riksspråk har gemensamt. Alla dialekter har ett ordförråd som består av dels särddialektala ord, dels ord som förekommer i både dialekter och riksspråk, men där orden vanligen uppvisar större eller mindre skillnader i fråga om semantik, fonologi, morfologi etc. Riksspråksgemensamma ord har inte alltid varit självklara beståndsdelar i svenska dialektordböcker. I de tidigaste ordböckerna, utkomna från 1700-talet och framåt, var urvalet begränsat till det särddialektala ordförrådet.² Den typen av urval hängde samman med dåtidens romantiska syn på forntiden, där dialekterna sågs som språkliga fornlämningar. Idén med dialektordböckerna var att tydliggöra de språkhistoriska banden bakåt i tiden och påvisa dialekterna som en källa till att berika riksspråkets vokabulär. Den första ordboken som avsåg att täcka in hela det dialektala ordförrådet i ett område var Ordbok öfver fryksdalsmålet utgiven 1878 av Adolf Noreen. Det fornnordiska inflytandet hade nu trätt i bakgrunden och Noreens junggrammatiska inriktning gör att alla ord ter sig lika intressanta som potentiella studieobjekt.

DL utkom drygt tio år före Noreens ordbok och Rietz uttrycker i ordbokens inledning en tidstypisk syn på dialekterna. Han framhåller vilka viktiga bidrag dialekterna kan ge för att upplysa om fornspråket och att de gör det möjligt för "riksmålet att utur allmogemålens ordförråd rikta sig" (s. II). Ljunggren (1939) säger i sin översikt över svenska dialektordböcker att DL är "tillkommen efter samma urvalsprinciper och med samma syften och i stort sett samma teknik" (s. 26) som de ordböcker som kommit ut före den. Men man behöver inte bläddra länge i lexikonet för att se att Rietz definitivt

² För en översikt om utvecklingen av svenska dialektordböcker, se Ljunggren (1939).

inte uteslutit alla ord som är gemensamma med riksspråket och detta intryck bekräftas av honom själv i ordbokens inledning:

Om något ord, som förekommer i svenska riksmålet, i munarterna har ett mycket egendomligt uttal, eller genom sina former, kön eller böjningar visar den ålderdomliga prägeln, har jag ansett det vara tjenligt att upptaga detsamma (s. IV).

Han exemplifierar sina principer med två ord, dels *ajn*, *ejn*, dels *dur*, *dyr*. *Ajn*, *ejn* 'en' finns med för att "visa den nära frändskapen med de uråldriga tungorna" och han framhåller särskilt att tveljuden, dvs. diftongerna, är viktiga. *Dur*, *dyr* 'dörr' har inkluderats för att visa "huru troget folket bevarat det gamla språkarfvet" (Inledning, s. IV).

3. Undersökning

Rietz har uppenbarligen låtit en del riksspråkliga ord komma med och den bild av hans urvalsprinciper som förmedlats av Ljunggren torde förtjäna en nyansering. Jag har därför kartlagt vilka riksspråkliga ord Rietz inkluderat och därefter undersökt följande: Hur förhåller sig dessa ord till Rietz uppställda kriterier? Uppvisar de några specifika uttal, former, kön eller böjningar och i så fall i vilken utsträckning? Kan det finnas andra skäl, än de Rietz nämner, till att "riksmålsord" kommit med? och slutligen: Hur presenterar Rietz dessa ord? Låter han t.ex. läsaren märka att orden också finns i riksspråket? Undersökningen begränsar sig till bokstaven h-, som omfattar 1554 lemman.³

4. Artiklar med beteckningen "Riksspr."

I 52 av artiklarna på h- visar det sig att Rietz själv meddelar att en av betydelseerna (eller den enda betydelsen) är riksspråklig, vanligen genom tillägg

³ Samtliga ord återfinns som huvud- eller sublemman inom partiet H (s. 233-289). Ord på h- förekommer även på andra platser i DL, men ingår inte i undersökningen. Även ord som inleds med <hv> och idag skrivs med <v> är inkluderade. Hänvisningsartiklar är ej medräknade.

av förkortningen ”Riksspr.” efter betydelseangivelsen. Det rör sig om nedanstående ord (redovisade i alfabetisk ordning och åtskilda med tankstreck). För varje ord anges ordklass eller genus samt betydelser i starkt reducerad form. Betydelsen är ersatt med ”=” i de fall den är (nästan) identisk med uppslagsordet och latin är översatt till svenska. Förekomsten av vissa för undersökningen relevanta uppgifter visas i form av förkortningar, där **R.** står för ”Riksspr.” eller motsvarande, **U.** uttalsuppgifter, **B.** böjningsuppgifter och *SF* markerar sidoformer till lemmat.⁴

hak n. ingående vinkel eller hörn **R.** – **hake** **1** m. 1) = **R.** 2) bössa – **hakka** v. 1) smått sönderdela; begabba **R.** 2) kälta 3) hosta oupphörligt 4) stamma 5) icke kunna läsa rent – **hakkspett** m. 1) = **R. U.** 2) lång och mager människa – **hals** m. 1) = **R. 2)** förhus, smal ingång 3) näthörn – **ham** **2** m. knäveck [anv. i riksspråkliga uttryck följer] – **hammar** **2 B.** m. 1) hammare **R.** 2) nacken på yxor el. likn. 3) ett stycke trä till sidslå i en skrinda eller stege – **hamn** **2** m. 1) = **R. 2)** stycke mellan sömmar – **hamna** v. 1) inlöpa i hamn **R.** 2) gå in med båten baklänges till hamnen 3) ro baklänges – **hampa** f. = **R.** – **hand** f. 1) = **R. U. B.** 2) ett gifvet quantum lin – **hane** m. tupp **R.** – **hank** m. 1) det hvarmed något hänges **R. U.** 2) handtag på byttor o.d. 3) ok 4) sölf i en väf **U.** – **hare** m. = **R. U.** – **harva** v. 1) = **R. U.** 2) räfsa **U.** – **hastiger** *SF* adj. 1) = **R. 2)** lättretlig **U.** – **hatt** **2** m. 1) betäckning för öfre delen af en sak **R.** 2) sädesskyl – **hatta** v. 1) betäcka (en sädesskyl) med hatt **R. 2)** uppsätta sädesskylar – **hejda** v. 1) återhålla **R.** 2) vara njugg – **helg** f. 1) helg, högtid **R. 2)** marknad – **hiskeligt** adj. 1) rysligt **R. 2)** ganska mycket – **hit** interj. 1) adv. hit **R. U.** 2) utrop till dragare – **hjerta** n. = **R. U. B.** – **ho** **2** m. 1) = **R. 2)** vattenränna – **hólme** m. **B.** 1) = **R. 2)** backe i en kärräng 3) ett mindre åkerfält 4) en mindre äng – **hop** **1)** m. samlad mängd **R. U.** 2) præp. hop, ihop **R.** – **hóppa** v. hoppa **R. U.** – **hora** f. 1) dam i kortleken 2) = **R.** – **horn** n. 1) = **R. U.** 2) udden af ett näs 3) träet, i hvilket räfs-pinnarne äro slagne – **hov** n. vett, måtta **R.** – **hud** f. = **R.** – **huka** v. 1) sitta på huk **R. U.** 2) aftaga; om månen – **huld** adj. välvillig **R. U.** – **hult** *SF* n. 1) skogsbacke, skogsdunge **R. 2)** hög backe – **hum** m. misstanke, dunkelt begrepp **R.** – **hund** m. = **R. U.** – **hus** n. 1) = **R. U. B.** 2) rum i en byggnad – **huttla** v. **R.** 1) narra, bedraga 2) vara senfärdig och förtretlig – **huv** m. 1) litet rundt tak **R. 2)** öfverbyggnad öfver en gödselstad 3) gödselrum – **hvila** *sej* v. 1) = **R. 2)** gå till den sista vilan – **hvina** **B.** v. 1) = **R. U. B.** 2) gråta, jämra sig **U. B.** – **hvit** adj. = **R. U. B.** – **hår(d)** adj. 1) = **R. 2)** sur, syrlig. **U.** – **häfta** v. 1a) löst fästa **R.** b) sätta hälla på (en häst) c)

⁴ Artiklarna kan även innehålla uppgifter om fornspråk, variantformer, sammansättningar, konstruktioner m.m. Talesätt och språkprov förekommer. Alla proveniensangivelser har utelämnats.

hindra **U.** 2a) vara hindringsamt b) dröja, vänta – **hälla** v. 1) luta emot **R.** 2) förflyta 3) närma sig, lacka åt – **hänglås** n. 1) ett slags rörligt lås **R.** 2) efterhängsen människa 3) våp; drumlig människa – **häpen** adj. 1) **R.** bestört 2) skyndsamt och på något begärlig – **här** adv. = **R.** – **häst** 2 m. = **R. U.** – **hätta** 3 **B.** f. [1]) karlmössa **R. U.** 2) fruntimmersmössa 3) få något i hättan, få ett rus – **höffsa** v. 1) hyfsa, städa **R.** 2) **höffsa öpp** slå, basa – **höna** f. 1) = **R.** 2) enfaldig människa.

För 37 av de 52 ovanstående orden anges särdialektala betydelser vid sidan av den riksspråkliga. Som synes redovisas den riksspråkliga betydelsen så gott som alltid först, i ett eget moment. Uttalsuppgifter⁵ förekommer relativt ofta och meddelas framför allt i anslutning till den riksspråkliga betydelsen. Avvikande genus omnämns av Rietz för ordet *hum*. Böjningsuppgifter anges sporadiskt och kortfattat för i huvudsak substantiv. För det enda starka verbet, *hvina*, redovisas dock temaformer utförligt.

5. Fler riksspråkliga ord

Det är lätt att upptäcka en hel del andra ord i DL som rimligen måste ha ingått i riksspråket på Rietz tid, utan att han explicit påtalat det. För att kunna avgöra vilka ord som bör ha varit en del av riksspråket har jag använt A. F. Dalins Ordbok öfver svenska språket som måttstock. Den kom ut 1850–53, alltså ungefär ett decennium före Rietz lexikon. Som riksspråkliga ord i DL har räknats alla etymologiskt identiska ord⁶ som föreligger också i Dalins ordbok och som han inte betecknat som populära⁷. Utfallet av jämförelsen är följande 160 lemman i DL⁸:

gvólv n. hvalf – **haalk** v. **B.** 1) halka 2) smyga sig 3) med möda och försigtighet dra sig fram – **haark** **B.** v. kratta bort – **hag** n. 1) gärdsgård i skogsmark och kring

⁵ Någon regelrätt ljudskrift är det aldrig fråga om; Rietz utgår från det vanliga alfabetets tecken med vissa tillägg. Se vidare DL förord s. IV och inledning s. V.

⁶ I de fall det vållat problem att avgöra om ett ord hos Rietz och Dalin har samma etymologiska ursprung, har SAOB konsulterats, någon gång även Hellquist. Om osäkerheten därefter kvarstått, har ordet fått utgå.

⁷ Dalin använder förkortningen ”pop.” för ”populärt, brukligt ibland det lägre folket” (s. 20). Med ”pop.” markerade ord på h-, vilka också upptas i DL, är: *handslög* adj., *hanka* v., *hasa* f., *hira* v., *hyde* n., *hyppja* v., *hönsa* v.

⁸ Tre ord i uppställningen inleds inte med h-, eftersom Rietz hämtat uppslagsorden från dialekter med initialt h-bortfall eller med uttalet /gv/ av äldre hv.

svedjeland **U. B.** 2) inhägnad af qvistar med snaror hvori foglar och harar fångas 3) flera näts sammansättning – **hagga B.** v. hugga **U. B.** – **hagrü** m. hafre **U.** – **hajt** f. stor, obebyggd skog. **U.** hed, sandhed, sandig slätt. **U.** hed. **U.** högländt sandmo bevuxen med skog – **hajta SF B.** v. heta, kallas **U.** – **hake 2** m. tapper man – **haku SF** f. haka – **hal** adj. 1) hård **U. B.** 2) hårdhänt 3) hårdsinad **U.** – **halku(g)** adj. något hal **U. B.** – **hallda B.** v. 1) hålla **U. B.** 2) hålla för, anse 3) väga 4) [i talesätt] – **halldfaster** adj. snål – **halver** adj. half **U.** – **ham 1** m. skugga, vålnad – **hamsa** v. 1) fingra på 2) hafsa; prata osammanhängande 3) leka med en katt 4) [i talesätt] **U.** – **hamsig SF** adj. pratande hams **U.** – **han** pron. = **B. U.** – **handfallen** adj. händig – **handkläde** n. hvit näsduk – **handsko** m. handske – **handspik** m. liten stör, hvarmed timmerstockarne bäras – **handsöl** n. 1) handpenningar 2) köpskål 3) drickspenningar – **hanekåmm** m. tuppkamm; så kallas blomman aurikel – **hanklä SF** n. handske **U.** – **hannka** v. varpa in (t.ex. tåg, not) – **har 1** n. hår **U.** – **harefötter** m. pl. 1) kattfot 2) bergsyra – **harhjärtä** adj. räddlifvad – **harka** f. kratta **U. B.** – **harpa** f. [1)] gammal, ful kärring 2) kärring, som pratar mycket 3) trollpacka – **harr SF** m. herre – **harull** m. ängsull – **harv B.** m. rifva, räfsa till hö – **harä** n. mellangärdet i ett kreatur – **has 1 B.** m. 1) knäveck 2) smalben 3) ben – **hasa B.** v. afskära senorna i ett djurs bakben – **haug** m. hög – **haugar B.** adj. hög **U. B.** – **haugmèltur** adj. 1) som talar högt 2) skrytaktig – **hauk SF** m. hök **U.** – **haust** m. höst **U.** – **he(dn)ing** m. 1) odöpt barn 2) [i uttryck] – **heim** n. 1) hem 2) vistelseort, boningsställe **U.** – **heima SF** adv. hemma **U.** – **heimat SF** adv. hemåt **U.** – **heiter SF** adj. het **U.** – **hejl** adj. hel, sammanhängande – **hejn SF** f. liten fin brynsten **U.** – **hel SF** f.? död – **hel(g)smess SF** f. allhelgonadag **U.** – **hel(g)smål SF** n. aftonringning **U.** – **helika SF** f. 1) hufvudbonad för fruntimmer **U. B.** 2) liten barnmössa **U.** – **helvite** n. = – **hemfö(d)ing B.** m. 1) hemmavarande vuxet barn 2) en som jemt varit i sina föräldrars hus **U.** – **hemlig** adj. 1) hemmastadd **U.** 2) bekant, förtrolig **U.** – **hemlighet** f. könsdelar – **hemsker SF** adj. 1) något förskräckt, försagd **U.** 2) fånig 3) illamående, matt – **hemsöka B.** v. besöka, helsa på – **henta SF** v. hemta, föra hem – **here** m. gosse, ung bonddräng **U.** – **herredag** m. riksdag – **hi(d)e** n. 1) ide **U. B.** 2) usel koja, illa beryktadt ställe **U.** 3) moras – **hink** m. en stolpe med häfstång m.m., hvarmed vatten upphemtas ur brunnen **U.** – **hinn SF** m. den onde – **hinna B.** v. = **U. B.** – **hira SF** f. förvirring **U.** – **hittug** adj. 1) fyndig 2) fintlig, förslagen – **hiva** v. 1) slänga, kasta **U. B.** 2) hissa upp **U.** – **hjelpa B.** v. = **U. B.** – **hjer SF** adv. här **U.** – **hjol** n. hjul **U.** – **hjolnål** f. järnpinne, som håller hjulet fast vid axeln – **hjon SF** n. människa **U.** – **hjul B.** v. trilla som ett hjul – **hjädan** adv. härifrån **U.** – **hjäörn** n. hörn – **ho SF** pron. hon **U. B.** – **hoegär SF** n. hufvudgård **U.** – **hógas** v. 1) trifvas 2) icke vara mörkrädd 3) längta efter **U. B.** – **hólländare SF** m. 1) arrendator af ett mejeri 2) brödknif – **hómmal** f. humla – **hóninggräs** n. gulmåra – **hornig** adj. full med finnar i ansigtet – **hórsafibla B.** f. 1) hökfibblor 2) slättergubbe – **hórsagök** m. 1) enkelbeckasin **U.** 2) en som talar eller sjunger hest – **hórskónung** m. tulkört – **hove** n. hufvud **U. B.** – **hovriddare**

m. brudsvan – **hu(g)** m. håg, minne – **hugg** n. = ; tillfälle – **huggulag** n. hufvudlag på en häst U. – **hugslaus** adj. minneslös – **hulld** n. hull – **hullung** m. hulling (på fiskkrok) – **humlekóppa** B. f. humleknopp – **hundfatt** adj. 1) skamlös; ofantlig 2) adv. oförskämtdt, orimligt – **hundfötter** pl. tät kvastfibbla – **hundsk** adj. skamlös – **hurra** v. 1) snurra i kring 2) hvina 3) dona, bullra U. B. – **hva?** [ordklass uppges ej] huru? – **hvaiv** SF B. v. 1) svänga, vifta 2) kasta, slänga – **hvarv** n. omgång U. – **hvaske** conj. hvarken U. – **hveila** f. stund; timme – **hveiv** f. 1) svängning 2) vef – **hvirred** adj. oredig U. – **hvitkulla** f. prästkrage – **hvitmat** m. mjölkmat – **hvo** adv. hvad? huru? – **hvoda** adv. hvarifrån – **hvorske** adv. hvarken, eller – **hvävla** B. v. = U. B. – **hwäsa** v. andas tungt U. – **hvättja** B. v. hvässa, bryna U. B. – **hybbele** n. ruckel – **hyddä** f. murad nisch i väggen nära spisen – **hygga sej** v. 1) förbättras; blifva ordentligare U. 2) kläda sig med omsorg – **hyla** v. 1) tjuta U. B. 2) hastigt rygga tillbaka 3) rysa af köld eller illamående – **hyll** m. fläder U. – **hyns** B. n. pl. höns – **hysing** B. m. rep, snodd U. – **hyssa** v. kasta med lätthet – **håi** n. hö U. – **hå-ken** m. hin onde – **håI** 1 adj. 1) urhålkad 2) ihålig 3) djup, tjock 3) [sic] uthungrig – **hålk** B. m. något urhålkadt 1) urhålkadt kärll 2) bladet eller röret, som omsluter axet eller blomstjelen 3) träkrans kring en brunn 4) källa i en träskartad omgivning 5) storätare 6) tjock pojke 7) drinkare – **håll** n. 1) styrka, varaktighet 2) hejd 3) syslande – **hållen** SF adj. 1) ansedd 2) underhållen – **hållning** f. underhåll af ett hemman – **hårkrank** B. m. ett slags stor mygga U. – **hårpös** m. harts U. – **hä(d)a** adv. hädan U. – **häfta** f. 1) förstoppning 2) urinstämna – **hägda** v. 1) spara U. 2) hålla måtta – **hägna** f. (inhägnad mark för) bete; betesmark – **häkkel** f. 1) linhäkla 2) tadel U. – **häkkelmakar** m. 1) en som gör häklor 2) tadelare – **håkkäl** v. 1) häkla U. 2) tadel U. – **häkta** v. 1a) dröja b) ej räcka till 2) hindra, emotstå – **hållning på hållning**, adv. nära slutet – **hånduger** adj. händig – **hänga** B. v. 1) = U. B. 2) vara något sjuk – **härberge** n. 1) handkammare U. 2) gästkammare U. – **hård** B. f. skuldra U. B. – **härma** v. = – **härnäst** adv. senast, sist – **häsja** SF f. torkställning U. – **hässja** SF v. flåsa häftigt U. B. – **håstfibler** f. pl. slåttergubbe – **håstgröning** f. skogsfråken – **häva** B. v. 1) häfva, vråka U. B. 2) [i talesätt] 3) hävas gåsa; om deg U. – **höfrö** n. ormrot – **högti(d)** SF m. 1) = 2) lappri – **hönsing** f. traktering af en nykommen person åt sina vänner – **hövla** v. 1) hyfla U. 2) banna upp – **hövvel** B. m. hyfvel – **inka** v. uppfordra vatten med [hink] – **årsträ** n. hårstrå.

Vad har fått Rietz att ta med dessa ord? Hur förhåller de sig till hans formulerade kriterier om uttal, former, kön och böjningar? Och varför har de inte fått beteckningen ”Riksspr.”?

Studerar vi inledningsvis kriteriet *uttal* kan sådana uppgifter uppträda på flera platser i artiklarna: i lemmat, i sidoförmer till lemmat (markerade med SF) och inuti artikeln. Eftersom Rietz princip är att välja den av honom för-

modade ”äldsta” dialektformen som lemma, får det till följd att dialektala uttal ofta kommer till synes där. Ser vi närmare på vilka uttalsdrag som förekommer i lemmana, rör det sig i 16 av orden om det som Rietz lyfte fram i sin inledning: diftonger och då alltid primära sådana. Relativt vanliga är även apokoperade former (*haalk, haark, hjul, hómml, häkkel* f., *häkkäl* v.) och ord med bevarat *ld* (*hallda, halldfaster, hulld*). För övrigt uppvisar orden en ganska stor variation av drag. När det gäller kriteriet *kön* är Rietz och Dalin oftast överens om genus hos substantiven. Skiljaktiga genus anger de för orden *harull, harv, harä, hejn, hoegär, hór safibla, hveiv, hägn, härd, hästfibler, hästgröning*. Ytterligare ett av Rietz kriterier är *böjning*. För samtliga åtta verb med stark böjning redovisas temaformer utförligt från olika platser. Även de svaga verben har i drygt hälften av fallen försetts med böjningsformer, men långt mindre detaljerade. För de två enda personliga pronomina, *han* och *ho*, anges kasusformer. Av adjektiven har 3 av 22 uppgifter om böjning. Vid 17 av 93 substantiv anges böjning, men då i regel bara enstaka uppgifter om formen i bf. sing. eller obf. plur. När det gäller urvalskriteriet *former*, ger Rietz inget besked om vad han avser med ”former”. Det kan inte gärna stå för uttals- eller böjningsformer, eftersom de omnämns särskilt. Det jag håller för mest troligt är att han menar ord som i dialekterna har andra suffix än i riksspråket, vare sig det gäller ordbildning eller grammatik. Till denna grupp hör framför allt adjektiv med ändelserna *-ug, -ot (-ed)* istället för *-ig* eller den gamla nominativändelsen *-er (-ur, -ar)*.

6. Varför inte ”Riksspr.”?

De just undersökta 160 orden har inte försetts med beteckningen ”Riksspr.”. Som torde framgå av uppställningen ovan, beror det inte på att riksspråkliga betydelser saknas. En jämförelse av betydelseuppgifterna för dessa ord i DL och Dalins ordbok visar att det finns betydelser som motsvarar varandra eller som tangerar eller innefattas i varandra, men också betydelser som saknar motsvarigheter. Betydelser med direkt motsvarighet hos Dalin återfinns för 97 av orden i DL. De har markerats i uppställningen med en understrykning under uppslagsordet. Man kan lägga märke till att en relativt stor andel av dem har ett lemma som återspeglar ett dialektalt uttal eller en dialektal form.

Det förekommer mycket sällan för de 52 orden med ”Riksspr.”. Sannolikt har de dialektala lemmarna starkt bidragit till att ”Riksspr.” inte varit aktuellt att använda. Det är inte *hajta*, *halver* och *hyns* som tillhör riksspråket, det är *heta*, *halv* och *höns*. Beteckningen ”Riksspr.” verkar alltså i praktiken ha stått för ”riksspråklig betydelse hos ett ord med riksspråkligt lemma”.

Men vad gäller för de ord som *inte* har ett dialektalt lemma – varför har Rietz inte använt ”Riksspr.” här? Jag kan peka på två tänkbara orsaker. Den första är att orden, trots att de finns med hos Dalin och inte markerats som populära, inte uppfattats som riksspråkliga av Rietz. Efter en kontroll av bruklighetsangivelser i SAOB och geografiska utbredningsuppgifter i dialektsamlingarna vid Dialekt- och folkminnesarkivet i Uppsala kan följande tio ord anses dialektala eller provinsiella: *hannka v.*, *harka f.*, *hel(g)smess*, *helika*, *hybbele*, *hyll*, *häfta*, *härd*, *hässja* och *hästgröning*.⁹ Den andra orsaken till att en betydelse inte betecknats med ”Riksspr.” kan vara att ordet inte tillhört det *allmänna* riksspråket. Några av orden har förhållandevis speciella betydelser och kan inte ha använts särskilt frekvent. Flera av växtbenämningarna hör dit, men också ord som *hasa*, *herredag*, *hovriddare*, *humleköppa*, *hvättja*, *hysing* och *häckelmakar*.

7. Slutord

Undersökningen av orden på h- har visat att 52 ord har en betydelse som Rietz betecknat som riksspråklig. Därutöver har ytterligare 160 ord visat sig föreligga i Dalins ordbok. För båda grupperna framstår dialektalt uttal som det viktigaste av Rietz urvalskriterier. Övriga kriterier är representerade i mer begränsad omfattning, med undantag för böjningar av starka verb. Vanligt förekommande är mer eller mindre särdialektala betydelser. Ibland tas de upp vid sidan av en riksspråklig betydelse, andra gånger utgör de hela betydelseinnehållet. Sannolikt har de särdialektala betydelserna spelat en betydande roll vid urvalet, men det är inte ett kriterium Rietz nämner. Även annat som presenteras i artiklarna, vilket inte studerats här, såsom ordens förekomst i sammansättningar och talesätt eller deras fornspråkliga historia

⁹ Som redan konstaterats har *härd* och *hästgröning* skiljaktiga genus hos Dalin och i DL, vilket kan ha varit bidragande till att Rietz inkluderat dem.

kan tänkas ha bidragit till att orden inkluderats. Till detta återkommer jag i mina fortsatta studier av DL.

Litteratur

Ordböcker

Dalin, A. F. (1850–53): *Ordbok öfver svenska språket*. 1–2. Stockholm.

DL, se Rietz.

Hellquist = Hellquist, Elof (1948): *Svensk etymologisk ordbok*. 1–2. 3. uppl.

Lund: Gleerup.

Noreen, Adolf (1878): *Ordbok öfver fryksdalsmålet: samt en ordlista från Värmlands älfdal*. Uppsala.

OSD = *Ordbok över Sveriges dialekter* (1991–2000). Häfte 1–3. Uppsala.

Rietz, Johan Ernst (1862–67): *Svenskt dialekt-lexikon eller Ordbok öfver svenska allmogespråket*. Lund. Register och rättelser av Erik Abrahamson, 1955. Uppsala: Lundequistska bokh.

SAOB = *Ordbok över svenska språket* (1898–). 1–. Lund.

Annan litteratur och källor

Carlquist, Gunnar (1948): *Lunds stifts herdaminne: från reformationen till nyaste tid*. Ser. II. Biografier. 2. Oxie kontrakt. Lund: Gleerup.

Dialektsamlingarna vid Dialekt- och folkminnesarkivet i Uppsala.

Ekholm, Gulli (2001): Johan Ernst Rietz – en lundensisk språkforskare. I: *Personhistorisk tidskrift* 2001:1, 57-65.

Ljunggren, Karl Gustav (1939): Svenska dialektordböcker. En översikt. I: *Vetenskaps-societeten i Lund. Årsbok* 1938, 21-37.

Noreen, Adolf (1903): *Vårt språk: nysvensk grammatik i utförlig framställning*. Första bandet. Lund: Gleerup.

Reinhammar, Vidar (1998): Om J. E. Rietz och Svenskt dialekt-lexikon. I: *Svenska landsmål och svenskt folkliv* 1997, 7-57.

Svenskt biografiskt lexikon. Artikeln J. Ernst Rietz. <sok.riksarkivet.se/SBL/Presentation.aspx?id=6753> (september 2015).

Eva Thelin
forskningsarkivarie, fil. kand.
Dialekt- och folkminnesarkivet
Institutet för språk och folkminnen
von Kræmers allé 21
SE-751 04 Uppsala
eva.thelin@sprakochfolkminnen.se

Vilka dialektord kvalar in i en standardspråklig ordlista?

Lena Wenner

This paper describes the work behind deciding which dialect words should be included in the Swedish Academy Glossary (SAOL) 14. All 500 dialect words in the glossary of SAOL 13 were examined and presented in 10 Internet-based questionnaires containing 50 words each. For each word, informants were asked if they knew it and whether they used it themselves. As a result, it is suggested that 100 words should be deleted; either because they are too dialectal and only have a limited geographical distribution, or because they nowadays are part of Standard Swedish. In addition, it is suggested that 14 dialect words that have not previously been included, should be incorporated in the glossary.

1. Inledning

Inför utgivningen av Svenska Akademiens ordlista 2015 (SAOL 14) fick jag i uppdrag av redaktionen att närmare undersöka vilka dialektord som skulle kunna ingå i ordlistan. I den här artikeln beskriver jag mitt arbete, diskuterar problem jag stötte på och redovisar resultatet av en enkätundersökning som jag genomförde.

2. Dialektord i riksspråksordböcker

Det finns en lång tradition att redovisa ett urval dialektord i riksspråksordböcker (se särsk. Thelander 1994:365f.; Ernby 2008:82; Mattsson 1974:72ff.). Man kan fråga sig vad syftet är med att förteckna ord som inte går att använda i hela språkområdet. Fyller sådana uppgifter någon funktion? När man läser i förorden till flera av riksspråksordböckerna kan man konstatera att redaktörerna anser att det i alla fall inte kan vara till skada för riksspråket att man tar med dialektord, så länge de är tydligt markerade. Tvärtom kan de vara en hjälp för ordboksanvändarna eftersom även dialekt-

ord kan förekomma i tryck.

Totalt finns det ca 500 dialektord i SAOL 13. Dessa har beteckningen ”prov.” som är en förkortning för *provinsialism*. I mitt arbete har jag utgått från det som Thelander (1994:369) beskriver som en *äkta provinsialism*. Det innebär att personer från en viss region utgår från att ett specifikt ord tillhör standardspråket, medan personer från andra regioner inte förstår varför ordet ska tas upp i en riksspråklig ordbok. (För en utförligare diskussion om vad en provinsialism är, se Thelander 1983.)

Inom språksociologin talar man ofta om fyra nivåer med neutralt standardspråk respektive lokal (eller genuin) dialekt som ytterligheter (se Thelander 1994:368; Andersson 1985:62). Av Tabell 1 framgår att vi på nivå 1 finner de ord som tillhör det neutrala standardspråket. Dessa ord står utan markering i SAOL. På nivå 2 återfinns ord som tillhör det regionala standardspråket och det är där de provinsiella orden i SAOL hör hemma. På nivå 3 och 4 finner vi regionala och lokala dialektord som snarare hör hemma i dialektordböcker än i en standardspråklig ordlista. I realiteten är det inte alltid enkelt att avgöra var gränserna mellan de olika nivåerna går och vissa ord som i dag finns i SAOL borde kanske i framtiden höra hemma i en dialektordbok och vice versa.

	Nivå 1	Nivå 2	Nivå 3	Nivå 4
Sociolingvistisk typ	neutralt standardspråk	regionalt standardspråk	regional dialekt	lokal dialekt
Markering i SAOL	ingen markering	prov.	ej med i SAOL	ej med i SAOL

Tabell 1: Orden i SAOL återfinns på språknivå 1 och 2.

Uppdraget innebar att stryka ord som inte längre kan anses vara provinsiella p.g.a. att de används så lokalt att de numera tillhör regional eller lokal dialekt (nivå 3 och 4 i Tabell 1). Tidigare provinsialismer kan också ha spritts till fler brukare och därmed övergått till att uppfattas som neutralt standardspråk. I dessa fall räcker det att etiketten ”prov.” stryks. Uppdraget innebar dessutom att föreslå nya provinsialismer. För det allmänna ordförrådet kan stora sökbara textdatabaser användas för att hitta högfrekventa ord som uppfyller kraven för ordlistan, t.ex. Språkbanken. En svårighet när man studerar

mer lokala ord är att de ofta saknas i den här typen av textdatabaser.

För att lösa den första uppgiften – att föreslå provinsialismer som ska strykas – behövde jag veta vilka provinsialismer som används i dag och hur geografiskt spridda de är. Här är det ett problem att det inte finns något modernt, lättillgängligt dialektmaterial att utgå från. Rietz dialektlexikon är utförligt, men utkom redan 1862–1867. Det finns också smärre artiklar om provinsialismer, men även de har några år på nacken (Bergman 1942, 1952; Levander 1927), liksom Institutet för språk och folkminnens digra dialektordsamlingar. Institutet har kontinuerligt registrerat dialekter i över 100 år, men de samlingar som är sökbara på ordnivå har inte uppdaterats sedan det på 1930-talet blev möjligt att göra ljudupptagningar. Det är därmed utifrån dessa ordsamlingar omöjligt att ta reda på vilka provinsialismer som används i dag och vilka ord som är föråldrade. För ordboksuppdraget behövde jag på ganska kort tid testa vilka dialektord som gemene man känner till och använder runt om i Sverige, och jag valde därför att göra en ny insamling.

3. Enkätstudie

Insamlingen genomfördes i form av en dialektundersökning via internet. Samtliga provinsialismer i SAOL 13 ingick i undersökningen. För att inte undersökningen skulle bli alltför krävande för informanterna fördelades orden på 9 webbenkäter med ca 50 ord vardera. Några veckor efter att undersökningen lanserats på internet tillfogades ytterligare en enkät till undersökningen med ord som inte ingår i SAOL 13, men som kunde vara tänkbara att införas som provinsialismer i SAOL 14. Dessa ord vaskades fram utifrån olika källor som redovisas i punkterna a–e nedan och som sedan sammanställdes i en 40 ord lång enkät.

a) 2006 sändes en uppskattad teveserie om dialekter. I samband med programmen uppmanades folk att bidra med dialektord som de uppfattar som typiska för sin region. Orden publicerades på SVT:s hemsida, sorterade region för region. Med hjälp av informanter från respektive region valdes gångbara provinsialismer ut.

b) Dialektfrågor som de senaste åren inkommit till Institutet för språk och folkminnen genomsöktes.

c) Språkvetare med arbetsplats i olika delar av landet ombads lämna förslag på provinsialismer.

d) Kommentarer efterfrågades i de 9 webbenkäterna, t.ex. om informanterna använder ett ord som de trots var standard men som de i ett senare skede upptäckte inte var bekant för andra personer i deras omgivning.

e) Dialektord som förekommer på internet genomsöktes, t.ex. Lingvistbloggen där Parkvall framställt utbredningskartor av ord som skiljer sig åt mellan olika regioner.

För varje ord i undersökningen får informanterna frågan om de känner till ordet och om de använder det själva eller inte. Termen ”provinsiella ord” är gissningsvis relativt okänd för folk i allmänhet. Därför efterfrågades istället om ordet är gammalt, dialektalt, standardspråkligt eller om det är ett slangord, för att få en idé om hur orden uppfattas.

Figur 1 visar en skärmdump av enkäten som innehåller ordet *abrovink*, med betydelsen ’listig lösning’. Informanten kan välja att enbart kryssa i om hen känner till ordet eller inte, men också ange om hen använder det själv. Det finns också ett fritextfält för frivilliga kommentarer. Enkätsvaren tyder på ett stort engagemang som tog sig uttryck i många kommentarer och reflektioner på varje ord.

Abrovink eller Abrovinch
Listig lösning

Jag känner inte till ordet
 Jag känner till ordet men använder det inte själv
 Jag använder själv ordet

Jag uppfattar ordet som:

Riksspråkligt
 Dialektalt
 Slang
 Ålderdomligt

Kommentar:

< Tillbaka Fortsätt >

Figur 1: Skärmdump som illustrerar enkätens utformning.

Kommentarerna handlar oftast om att folk anser att den definition som anges för ett visst dialektord är felaktig. Det är intressant eftersom betydelseangivelserna för orden i enkäterna är hämtade från SAOL. Detaljerade analyser av specifika ord skulle kunna visa om det är frågan om betydelseförskjutningar i fall där många är överens om en ny betydelse.

3.1 Avgränsningsproblem

Två frågor som dök upp under genomgången av provinsialismerna var vilka avgränsningar som borde göras och hur konsekvent urvalet av upplysningar bör vara för provinsialismerna.

Vanligast i SAOL är förstås att etiketten sätts på ordnivå, men även uttal och böjningsformer är markerade som provinsiella i vissa fall. De uttalsangivelser som anges som provinsiella är få och i några fall stämmer de inte överens med dagens uttal, exempelvis uttalas inte *knöka* längre med enbart ”öppet ö som i *göra*”, som det står i SAOL 13 (s. 627). Bland yngre västsvenskar hörs i dag även ett slutet ö-ljud i ordet *knöka*.

Det finns också ord i SAOL som kategoriseras som ”mest prov.” (alltså *mest provinsiella*), exempelvis *maja* ’pryda med grönt’, *rimpa* ’remsa’ och *skällning* ’tätning av fogar mellan tegelpannor med vitkalkat murbruk’. Det visar väl att det är svårt att dra gränsen för vad som är ett provinsialt ord, men antingen är det frågan om en provinsialism eller också får man anse att ordet är så spritt att det tillhör standardspråket.

När det gäller urval ska understrykas att SAOL inte är en definitionsordbok, men ordförklaringar finns av princip för mindre vanliga ord (Gellerstam 2009:20). Därför är det anmärkningsvärt att 18 provinsialismer saknar definition i SAOL, exempelvis *vedlider*, *lenväder* (’blidväder, töväder’) och *pyrighet* (’livlighet’). Anledningen till att de är markerade som ”provisiella” är ju att de inte förekommer över hela Sverige, och därför bör det betraktas som ännu viktigare att ange betydelsen för just de här orden, eftersom man per definition utgår från att det finns folk som inte känner till dem.

3.2 Föråldrade betydelseangivelser

Vid genomgång av kommentarerna till de ord som redan finns i SAOL vi-

sade det sig att informanterna hade svårigheter med en del av orden. Flera kommentarer gällde att de inte förstod den betydelse som angetts. Det är uppenbart att betydelseangivelserna i vissa fall blivit föråldrade, t.ex. förklaras *trave* med 'skyl'. Går man till uppslagsordet *skyl* i SAOL kan man läsa att det är 'en samling kärvar sammanställda för torkning på åkern'. Det är inte så ofta vi ser dessa kärvar nu för tiden, så det är självklart att både standardordet och det regionala ordet används ganska sällan. Andra exempel där betydelseangivelserna åldrats är 'vret' som förklarar provinsialismen *täkt* och 'klövsadel' som förklarar *klöv*.

En nackdel med skriftliga enkäter är att forskaren inte kan gripa in och förtydliga om det är uppenbart att en informant missuppfattat någonting. Relativt ofta uppger informanter att de använder ett visst ord, men av deras kommentarer inser man att de använder ordet i den standardsvenska betydelsen – inte i den regionala som anges i enkäten. Ett exempel är *kase*. Folk svarar att de känner till det i betydelsen 'rishög', vilket är den standardspråkliga betydelsen. Det verkar inte som att någon känner till *kase* i den provinsiella betydelsen 'gödselhög'. Ett annat exempel är *fåle* som folk känner till från Staffansvisan, och anger att det betyder 'häst'. Den provinsiella betydelsen 'föl' verkar vara relativt okänd. *Patt* känner folk inte till i den provinsiella betydelsen 'spene', bara den vardagliga betydelsen för 'bröst'. Några informanter tror också att de funnit fel i enkäten, t.ex. för ordet *kulting*. Det verkar självklart för informanterna att *kulting* är en liten gris, men i SAOL finns också den provinsiella betydelsen 'liten pojke'. Här verkar alltså de provinsiella betydelseerna vara på väg att försvinna.

3.3 Fördelar och nackdelar med webbundersökningen

Denna typ av insamling har många fördelar. Undersökningen publicerades på institutets hemsida och på dess Facebooksida. Inom två dagar hade 1 200 människor sett inlägget och inte mindre än 220 personer hade svarat på minst en av de tio enkäterna. Det är med andra ord ett enkelt sätt att snabbt få en hel del information. Jag når människor som jag inte känner och det gör materialet mer balanserat när det gäller faktorer som ålder samt geografisk och social tillhörighet. Det är möjligt att de personer som väljer att fylla i enkäterna är mer intresserade av språkfrågor än genomsnittet, men det bör inte ha någon betydelse för resultatet av denna studie.

Internetbaserade enkäter är ett bra sätt att samla in stora mängder data, men naturligtvis finns det även nackdelar. En nackdel är av rent teknisk art. Enkäterna har inte fungerat fullt ut i alla webbläsare och i mobiltelefoner, vilket naturligtvis medför ett visst bortfall.

Det finns ett annat problem som är allvarigare och svårare att komma åt. Jag har vänner som pratar mycket bred dialekt. Jag bad en av dem att fylla i ett par av enkäterna. När jag gick igenom svaren blev jag förvånad över att hon för många ord hade markerat att hon inte känner till dem, även om jag vet att hon använder dem dagligen. Ett ord som finns med i SAOL är *laddor* som jag vet att min vän ofta använder. Ändå svarade hon i enkäten att hon inte vet vad ordet *laddor* betyder. Det kan verka förvånande, men förklaringen är att det i enkäten – precis som i ordlistan – står *laddor* med definitionen ”Sko eller innerpjäxa av filt”. Min vän säger inte *laddor*, utan *laddera* och menar då sina tofflor. Hon förstod inte att *laddor* är samma sak som det hon kallar *laddera*. Det är inte heller konstigt. För det första skiljer sig betydelseerna åt och för det andra ser man sällan dialektord i skriftlig form, så det kan vara svårt att koppla uttalet till ett ord som man för första gången ser i skrift. Ett alternativ kunde vara att koppla en ljudfil till enkäten så att man också kunde lyssna på dialektorden, men det är inte helt oproblematiskt eftersom samma ord kan uttalas ganska olika i olika delar av landet. En del av avvikelserna mellan uttal och skrift fångas i kommentarerna, men exemplet ovan visar hur viktigt det är att också utföra djupintervjuer med några personer från varje dialektområde.

4. Med facit i hand – hur blev det till slut?

Totalt 800 enkäter besvarades. Utifrån enkätsvaren föreslog jag 100 ord som skulle kunna utgå ur nästa upplaga och 16 ord som skulle kunna införas. Det var svårare än jag trodde att hitta provinsialismer som används i dag. Ibland har jag varit tveksam till om ordet i fråga snarare är slang, vardagligt eller om det rent av är alltför dialektalt. Jag valde ut 40 tänkbara kandidater (om urvalet se avsnitt 3) som låg som en egen enkät i undersökningen och besvarades av 82 personer. Jag har haft strängare tillämpning av kriterierna när det gäller användning och kännedom om de ord som skulle kunna kvala in i SAOL än för ord som redan finns i ordlistan.

Av de 100 ord jag föreslog kunde utgå ur ordlistan är det åtta ord som redaktionen valde att behålla. Dessa är: *bamba* 'barnbispisning i skola', *bräcka* 'brink', *färsig* 'som består av färs', *glad* 'öppet ställe, glänta', *glunt* 'yngling, pojke', *grisk* 'ivrig', *inkiett* 'otålig, ivrig' och *skällning* 'tätning av fogar mellan tegelpannor med vitkalkat murbruk'.

Exempel på avgränsningar som måste övervägas är hur föråldrade dialektord ska behandlas i SAOL. Det framgick av kommentarerna till enkäterna att vissa föråldrade dialektord lever kvar tack vare att de finns på pränt. Det är ingen som uppger att de använder *glunt* om *pojke*, men de känner naturligtvis till ordet från Gluntarna av Wennerberg även om de kanske inte vet vad ordet betyder. Det kan finnas en poäng i att ett ord som *glunt* finns med i ordlistan även om det inte är särskilt levande i språket. Det är ett ord som folk stöter på.

Sedan har vi *bräcka* som är bekant för många för att det ingår i namnet *Fjärås Bräcka*. Det finns också andra ord som folk känner till för att de ingår i ortnamn, t.ex. *vång* 'gärde, fält' och *flad* 'grund havsvik'. Namnet *Vånga* finns både i Skåne och i Östergötland. *Fladen* i Kattegatt känner folk till från sjörapporterna. Det framgår inte av enkäterna om informanterna känner till ordens egentliga betydelse. Min gissning är att appellativen i dessa fall har övergått till namnelement. Då kanske man snarare ska leta efter namnelementens betydelser i ett ortnamnslexikon än i SAOL. Samtidigt är det en fin service att dessa ord finns med och det är inte någon stor grupp ord, så de är inte särskilt utrymmeskrävande.

Av de ca 350 ord som jag föreslog skulle kvarstå har redaktionen tagit bort två. Det är *flane* 'dum person' och *skrövla* 'skrynkla'.

Av de nya ord jag föreslog togs 14 ord med, bl.a. *borrfluga* 'borrmaskin', *butta* 'knuffa', *exter* 'hyss, dumheter', *fresta* 'försöka', *katig* 'kavat, kaxig', *pumla* 'julgranskula' och *säckkärra* 'hopfällbar bagagevagn med två hjul'. Jag föreslog att ordet *bänglig* 'besvärlig' som jag uppfattar som norrländskt skulle införas. Det kom in i ordlistan, men inte som provinsialism, utan med etiketten "vardagligt". Vidare föreslog jag att det sydsvenska ordet *gräddbulle* 'chokladöverdragen skumboll' skulle tas med i ordlistan. *Gräddbulle* finns numera i SAOL, men det ansågs av redaktionen vara så vanligt förekommande att det inte klassades som en provinsialism.

5. Tankar inför framtiden

Nu är arbetet avslutat för den här gången. Vad som återstår att fundera över är vad som hade kunnat göras annorlunda för att uppnå säkrare resultat av vilka provinsialismer som används i dag. Enkäterna var i längsta laget och ibland uppstod tekniska problem. Informanterna fick välja vilken eller vilka av de 10 enkäterna de ville besvara, men majoriteten valde ändå att besvara enkäterna i nummerordning. Det var ett fåtal informanter som besvarade samtliga enkäter, vilket fick som följd att de enkäter som låg sist i undersökningen har betydligt färre svar. Ska man göra om undersökningen i framtiden kunde man tänka sig att orden kommer slumpvis och att försökspersonerna besvarar så många ord de vill.

Det försvann som sagt 100 provinsialismer medan bara ett tiotal nya infördes. Det är enklare att sortera bort gamla provinsialismer som inte används och som därmed inte har något berättigande i ordlistan än att föra in nya provinsialismer, men det senare är väl så viktigt för att ge stöd åt dagens språkanvändare. En metod som kan användas för att söka efter nya ord är att använda landets lokaltidningar mer. Genom att gå igenom vilka ord som förekommer i lokaltidningarna, men som inte tillhör standardspråket, skulle man få ett bra underlag för vilka ord som skulle ha sin plats i SAOL. Här kan tilläggas att de ord som funnits i någon av de tidigare ordlistorna är sökbara i databasen SAOLhist som är tillgänglig för alla.

Ett annat viktigt arbete framöver är att uppdatera arkivens dialektord-samlingar. I dag är det enkelt att spela in dialekttalande personer. Problemet är att det tar tid att transkribera eller excerpiera inspelningar, men det sker en utveckling på detta område. Det finns program och appar som omvandlar tal till text och i framtiden kan de förhoppningsvis fungera tillräckligt bra även på dialektmaterial. I väntan på detta är ett alternativ att samla in ljudmaterial via internet. Allmänheten kan ges möjligheten att skicka ljudfiler via en webbenkät där de uttalar vissa ord. Det skulle då också kunna finnas en karta där informanten kan pricka in var hen är uppväxt. Klickar informanten sedan på en region får hen höra olika uttal för ett och samma ord. Kartan kan göras ännu intressantare om den kompletteras med etymologier eller annan information om några av orden. Detta är förslag på ett sätt att uppdatera institutets ordsamlingar, vilket i framtiden kunde underlätta arbetet att revidera provinsialismerna i SAOL.

Litteratur

Ordböcker

- Rietz, Johan Ernst (1867): *Svenskt dialekt-lexikon. Ordbok öfver svenska allmoge-språket*. 1–2. Malmö, Köpenhamn, Leipzig & London [1862–] 1867. Fotolitografisk reproduktion. Lund 1962. 3. Register och rättelser av Erik Abrahamson. Uppsala & København 1955.
- SAOL 13 = *Svenska Akademiens ordlista över svenska språket*. (2006). 13 uppl. Stockholm: Norstedts.
- SAOL 14 = *Svenska Akademiens ordlista över svenska språket*. (2015). 14 uppl. Stockholm: Norstedts.

Annan litteratur

- Andersson, Lars-Gunnar (1985): *Fult språk*. Stockholm: Carlsson.
- Bergman, Gösta (1942): Sydliga provinsialismer i det bildade svenska talspråket. I: *Arkiv för nordisk filologi* 56, 125-250.
- Bergman, Gösta (1952): Nordliga provinsialismer i det bildade nordliga talspråket. I: *Arkiv för nordisk filologi* 66, 131-215.
- Ernby, Birgitta (2008): Svenska dialektord i riksspråkliga ordböcker. I: *Nog ordat? Festskrift till Sven-Göran Malmgren*. Göteborg: Meijerbergs institut för etymologisk forskning.
- Gellerstam, Martin (red.) (2009): *SAOL och tidens flykt: några nedslag i ordlistans historia*. Stockholm: Norstedt.
- Levander, Lars (1927): I vad mån beaktas uppsvenskt språkbruk vid Svenska Akademiens ordboksarbete? I: *Nysvenska studier* 7, 30-42.
- Lingvistbloggen = <lingvistbloggen.ling.su.se/?p=1042> (augusti 2015).
- Mattsson, Gösta (1974): Ordlistans tionde upplaga – riktlinjer och kommentarer. I: T. Johannisson & G. Mattsson: *Svenska Akademiens ordlista under 100 år*. (Skrifter utg. av Svenska språknämnden 55.) Stockholm: Esselte studium.
- SAOLhist = <spraakdata.gu.se/saolhist/> (augusti 2015).
- Thelander, Mats (1983): Provinsialismer i inom- och mellanregionalt perspektiv. I: *Språk och tradition. Festskrift till Sven Benson*. Uppsala: Dialekt- och folkminnesarkivet.

Thelander, Mats (1994): prov. utan värde? Om provinsialismer i ordböcker – och i verkligheten. I: *Språkbruk, grammatik och språkförändring. En festskrift till Ulf Teleman 13.4.1994*. Lund: Institutionen för nordiska språk.

Lena Wenner
forskningsarkivarie, fil.dr
Instituttet för språk och folkminnen
Vallgatan 22
SE-411 16 Göteborg
lena.wenner@sprakochfolkminnen.se

KONFERENZPROGRAMM

13. Konference om Leksikografi i Norden
København 19.-22. maj 2015

PROGRAM

Nordisk Forskningsinstitut, Københavns Universitet
Njalsgade 136

Tirsdag 19. maj 2015	
18.30-21.30	Lokale 27.2.23, Nordisk Forskningsinstitut, Njalsgade 136, 2. sal Uformelt velkomstarrangement med let buffet Registrering
20.00-21.00	Mulighed for rundvisning på Ordbog over det Norrøne Prosasprog (ONP) og Ømålsordbogen (ØMO) Rundvisning kl. 20.00 og 20.30

Onsdag 20. maj 2015		
8.30-	Lokale 27.0.47 Registrering	
9.00-9.30	Sal 1: 22.0.11 Officiel åbning og velkomst Morgensang	
9.30-10.30	Sal 1: 22.0.11 Plenarforelæsning 1 Maria Koptjevskaja Tamm, Stockholms universitet <i>Temperaturord: lexikal typologi och lexicografi</i>	
10.30-11.00	Kaffepause - foran 22.0.11	
	Sal 1: 22.0.11 Tema: søgning og struktur	Sal 2: 27.0.09 Leksikon: kollokationer
11.00-11.30	Seán Vrieland <i>Stamtræer og digitale etymologiske ordbøger</i>	Aurelija Griškevičienė & Sturla Berg-Olsen <i>Å gå seg vill i en ordbok: seg og V + seg i tospråklige ordbøker</i>
11.30-12.00	Kristín Bjarnadóttir <i>The case for normalization: Linking lexicographic resources for Icelandic</i>	Margunn Rauset <i>Når finst eit fast uttrykk?</i>

12.00-13.00	Frokost/Lunsj - kantinen i Bygning 23	
	Sal 1: 22.0.11 Tema: søgning og struktur	Sal 2: 27.0.09 Leksikon: kollokationer m.v.
13.00-13.30	Erik Bäckerud <i>Indexeringen av SAOB</i>	Torben Arboe <i>Konventionelle figurative enheder (idiomer mv.) i Jysk Ordbog og andre større ordbøger</i>
13.30-14.00	Erik Bäckerud <i>Indexeringen av SAOB: Softwaredemonstration</i>	Louise Hallstrøm Abildgaard & Amalie Glargaard & Carsten Hansen <i>Facebook – nyt brugersegment for Ordbog over Dansk Talesprog</i>
14.00-14.20	Sal 1: 22.0.11 Posterpræsentationer: Pär Nilsson <i>En kognitiv semantisk analys av SAOB:s formler för semantisk förändring</i> Erla Hallsteinsdóttir & Uwe Quasthoff & Sabine Fiedler <i>Frequency Dictionaries as a tool for analog and digital dictionary users</i> Simonetta Battista & Ellert Þór Jóhannsson <i>Ordbog over det norrøne prosasprog – fra seddelsamling til online-version</i> Louise Hallstrøm Abildgaard & Amalie Glargaard & Carsten Hansen <i>Ordbog over Dansk Talesprog (ODT) på Facebook (ODT-F)</i>	
14.20-14.50	Kaffepause - foran 22.0.11	
14.50-18.30	Udflugt - mødested: Njalsgade ud for Bygning 22	

Torsdag 21. maj 2015		
9.00-10.00	Sal 1: 22.0.11 Plenarforelæsning 2 (Tema) Esben Alfort, lingvist og forskningsleder, Ankiro, København <i>Ord og IT</i>	
	Sal 1: 22.0.11 Tema: søgning og struktur	Sal 2: 27.0.09 Lemmaselektion
10.00-10.30	Caroline Sandström & Carola Åkerlund <i>Personbetecknande verbavledningar i Ordbok över Finlands svenska folkmål</i>	Sturla Berg-Olsen & Daniel Gusfre Ims <i>Dansk og norsk i norske ordbøger</i>

10.30-11.00	Ellert Þór Jóhannsson & Simonetta Battista <i>Ordbog over det norrøne prosasprog online: Struktur og brug</i>	Lena Wenner <i>Vilka provinsiella ord kvalar in i en standardspråklig ordlista?</i>
11.00-11.30	Kaffepause - foran 22.0.11	
	Sal 1: 22.0.11 Tema: brugeradfærd	Sal 2: 27.0.09 Lemmaselektion
11.30-12.00	Poul Hansen <i>Hvordan har ordbogsbrugernes præference for forskellige emnespecifikke ordlister ændret sig i perioden 2005-2014?</i>	Eva Thelin <i>Rietz och riksspråket</i>
12.00-12.30	Tove Bjørneset <i>LEXIN i Norge: Hva sier brukerne?</i>	Louise Holmer & Sven-Göran Malmgren & Monica von Martens <i>SAOLhist.se – för allmänt och vetenskapligt bruk</i>
12.30-13.30	Frokost/Lunsj - kantinen i Bygning 23	
	Sal 1: 22.0.11 Tema: brugeradfærd og apps	Sal 2: 27.0.09 Lemmaselektion m.v.
13.30-14.00	Henrik Lorentzen & Nicolai Hartvig Sørensen & Lars Trap-Jensen <i>Den Danske Ordbog som app: Hvorfor og hvordan?</i>	Håkan Jansson <i>Från "tandborstord" till "memilord"</i>
14.00-14.30	Louise Holmer & Ann-Kristin Hult & Emma Sköldberg <i>SAOL-appen under luppen: Om användningen av Svenska Akademiens ordlista i appformat</i>	Carina Nilstun <i>NAOB: På vei mot en nasjonalordbok for bokmålet: Nyordprosjektet «2014 nyord i 2014»</i>
14.30-15.00	Henrik Køhler Simonsen <i>Den mobile ordbogsbruger: Ny vin på nye flasker</i>	Lars Brink <i>Synonymitis</i>
15.00-15.30	Kaffepause - foran 22.0.11	
	Sal 1: 22.0.11 Tema: brugeradfærd m.v.	Sal 2: 27.0.09 Leksikon: sprogteknologi
15.30-16.00	Marit Hovdenak & Ingunn Indrebø Ims <i>Bruk av digitale ordbøker i norsk skole</i>	Bolette Sandford Pedersen & Sussi Olsen & Sanni Nimb & Anna Braasch <i>Betydningsinventar – i ordbøger og i løbende tekst</i>

16.00-16.30	Anne-Line Graedler <i>Lærerstudenters bruk av digitale ordbøker i engelskundervisningen</i>	Sanni Nimb & Bolette Sandford Pedersen <i>Fra begrebsordbog til sprogteknologisk ressource: verber, semantiske roller og rammer – et pilotstudie</i>
16.30-17.00	Øystein Eek <i>Ordnett og den digitale ordbokbrukeren</i>	Halldóra Jónsdóttir & Þórdís Úlfarsdóttir <i>Processering af synonymkomponent i en flersproget ordbog</i>
19.00-23.30	Festmiddag Restaurant Bellahøj Bellahøjvej 20 www.bellahøj.dk	

Fredag 22. maj 2015		
9.00-10.00	Sal 1: 22.0.11 Plenarforelæsning 3 Oddrun Grønvik, hovedredaktør, Norsk Ordbok 2014, Oslo <i>Morsmålsleksikografien i ei skiftande verd</i>	
10.00-10.15	Kaffepause - foran 22.0.11	
	Sal 1: 22.0.11 Tema: nye brugere til gamle ordbøger	Sal 2: 27.0.09 Tema: fra base til bruger
10.15-10.45	Thomas Troelsgård & Nicolai Hartvig Sørensen <i>Nyt liv til støvede ordbøger</i>	Erla Hallsteinsdóttir & Uwe Quasthoff & Sabine Fiedler <i>Frekvensordbøger som værktøj for analoge og digitale ordbogsbrugere</i>
10.45-11.15	Christian-Emil Ore <i>Gamle ordbøger og den digitale ordboksbruger</i>	Øystein Eek & Espen Karlsen <i>Digitalt korpus og ordbok over norsk middelalderlatin (MIDLAT)</i>
11.15-12.30	Sal 1: 22.0.11 Generalforsamling Nordisk Forening for Leksikografi	
12.30-13.30	Frokost/Lunsj - kantinen i bygning 23	
13.30-14.30	Sal 1: 22.0.11 Plenarforelæsning 4 (Tema) Holger Hvelplund, direktør, IDM Danmark <i>Digital publicering af ordbøger – nye udfordringer, nye muligheder</i>	

	Sal 1: 22.0.11 Tema: brugergrupper	Sal 2: 27.0.09 Tema: fra base til bruger
14.30-15.00	Patrick Leroyer <i>Bruger- og ekspertinddragelse i konstruktionen af online ordbøger</i>	Helga Hilmsdóttir & Nina Martola <i>Flerspråkslexikografi och obesläktade språk: erfarenheter från finska Islex</i>
15.00-15.30	Kaffepause - foran 22.0.11	
	Sal 1: 22.0.11 Tema: brugergrupper	Sal 2: 27.0.09 Leksikon: domæner
15.30-16.00	Margrethe Heidemann Andersen & Jørgen Nørby Jensen & Anita Ågerup Jervelund <i>Fordele og ulemper ved løbende opdatering af den digitale Retskrivningsordbog</i>	Ruth Vatvedt Fjeld & Arash Saidi <i>Hva er egentlig et akademisk vokabular?</i>
16.00-16.30	Ken Farø <i>Nyere tilgange til leksikografi – med nedslag i moderne elektroniske og trykte opslagsressourcer</i>	Jon Helgason & Emma Sköldberg <i>Kodifiering av multietnisk ungdomsspråk</i>
16.30-17.00	Sal 1: 22.0.11 Afslutning	

